

# Argumentation Synthesis following Rhetorical Strategies

Henning Wachsmuth \*   Manfred Stede \*\*   Roxanne El Baff †

Khalid Al-Khatib †   Maria Skeppstedt †\*\*   Benno Stein †

\* Paderborn University, Paderborn, Germany, [henningw@upb.de](mailto:henningw@upb.de)

\*\* University of Potsdam, Potsdam, Germany, [stede@uni-potsdam.de](mailto:stede@uni-potsdam.de)

† Bauhaus-Universität Weimar, Weimar, Germany, [<first>{.<last>}+@uni-weimar.de](mailto:<first>{.<last>}+@uni-weimar.de)

‡ Linnaeus University, Växjö, Sweden, [maria.skeppstedt@lnu.se](mailto:maria.skeppstedt@lnu.se)

## Abstract

Persuasion is rarely achieved through a loose set of arguments alone. Rather, an effective delivery of arguments follows a rhetorical strategy, combining logical reasoning with appeals to ethics and emotion. We argue that such a strategy means to *select*, *arrange*, and *phrase* a set of argumentative discourse units. In this paper, we model rhetorical strategies for the computational synthesis of effective argumentation. In a study, we let 26 experts synthesize argumentative texts with different strategies for 10 topics. We find that the experts agree in the selection significantly more when following the same strategy. While the texts notably vary for different strategies, especially their arrangement remains stable. The results suggest that our model enables a strategical synthesis.

## 1 Introduction

The primary use of arguments in natural language is to persuade others of a stance towards a controversial topic (Mercier and Sperber, 2011). According to Johnson and Blair (2006), an argument is logically cogent if its premises are relevant as support for its conclusion, individually acceptable, and together sufficient to draw the conclusion. In real life, however, argumentation is by far not only about logic (Allwood, 2016). Without a *rhetorical strategy*, arguments will hardly ever unfold their persuasive effectiveness.

Following Aristotle (2007), we see a rhetorical strategy as the purposeful encoding of three means of persuasion in a well-arranged and well-phrased speech or text: *logos* (providing logically reasoned arguments), *ethos* (demonstrating good character and credibility), and *pathos* (evoking the right emotions). Listeners or readers then decode the encoding, forming their view of the author’s logos, ethos, and pathos. In the realm of the area of computational argumentation, rhetorical strategies are particularly relevant for technologies that synthesize argumentative text and that aim to deliver arguments effectively.

Existing argument mining research largely focuses on the logical structure of arguments, identifying their units (premises vs. conclusions) and relations (support vs. attack). Recently, a few studies have tackled strategy-related aspects, such as explicit expressions of ethos (Duthie et al., 2016) and the effects of logical and emotional arguments across audiences (Lukin et al., 2017). So far, however, strategies have not been considered in argumentation synthesis, which altogether has not received much attention (see Section 2).

In this paper, we study the role of rhetorical strategies when synthesizing argumentation. In particular, we consider monological argumentative texts where an author seeks to persuade target readers of his or her stance towards a given topic, such as news editorials and persuasive essays. Conceptually, we argue that an author synthesizes a text of such genres in three subsequent steps:

1. *Selecting content* in terms of argumentative discourse units (along with facts, anecdotes, and similar) that frame the given topic in a way that is effective for the intended stance,
2. *arranging the structure* of the units considering ordering preferences, and
3. *phrasing the style* of the resulting text to match the genre and the encoded means of persuasion.

We expect that the encoding of the means of persuasion becomes manifest in measurable text properties related to selection, arrangement, and phrasing. Figure 1 sketches the analysis of such properties in a

Topic: Does college edu help financial security? Stance: con

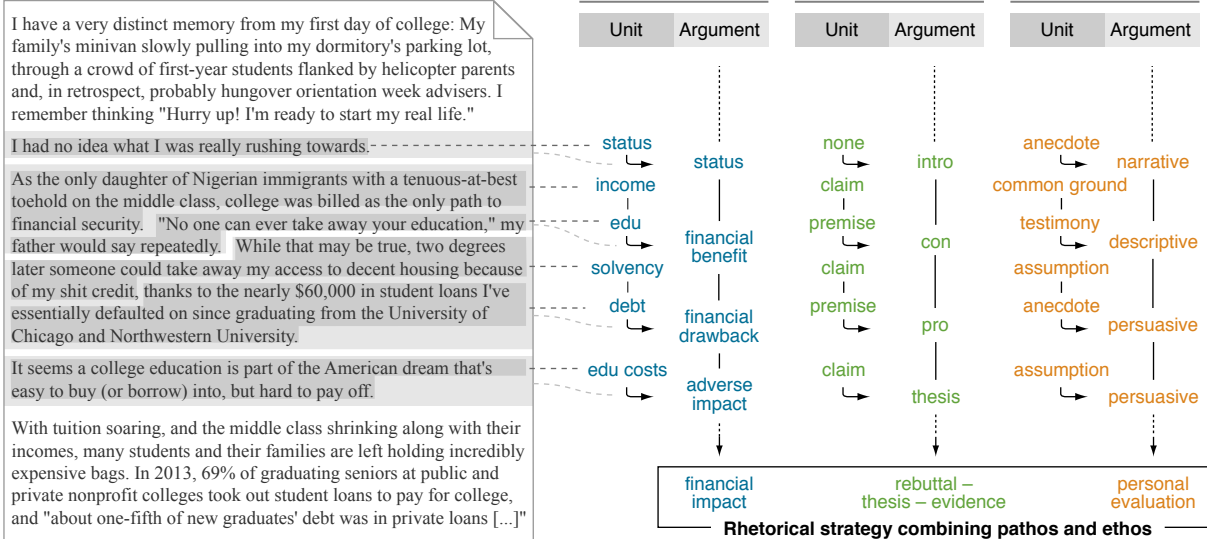


Figure 1: A monological argumentative text (a news editorial from Al-Khatib et al. (2016)), along with exemplary content, structure, and style properties captured by the envisioned model of rhetorical strategies.

pathos and ethos-oriented news editorial. Thereby, it also hints at our envisioned computational model of rhetorical strategies. We discuss this model and its application when synthesizing an argumentative text following a particular rhetorical strategy in Section 3. In Section 4, we design a dataset for studying argumentation synthesis with 200 argumentative discourse units for 10 pairs of topic and stance from the Arg-Microtexts corpus (Peldszus and Stede, 2016). To evaluate the adequacy of the general synthesis process reflected by our model, we then carry out an experiment where 26 human experts select, arrange, and re-phrase short argumentative texts following principled rhetorical strategies, thereby also creating benchmark data for computational approaches (Section 5).

We find that humans agree significantly more when selecting argumentative discourse units to encode specific means of persuasion, and we observe clear differences between strategies. At the same time, the arrangement of the texts remains stable across strategies. Their phrasing requires further investigation in future work. The results suggest that our model captures both general and strategy-specific aspects of synthesizing argumentation. As such, it defines a first substantial step towards computational argumentation synthesis following rhetorical strategies. In total, our main contributions are:

1. A *computational model* of rhetorical strategies for synthesizing monological argumentative texts.
2. A *benchmark dataset* with manually synthesized arguments that follow rhetorical strategies.
3. *Empirical evidence* that humans agree in strategy-specific argumentation synthesis.

## 2 Related Work

Aristotle (2007) revealed that persuasion is achieved through a well-arranged speech with a clear and appropriate style that delivers an adequate mix of the three means of persuasion: (1) *logos*, the use of logically reasoned arguments, (2) *ethos*, the demonstration of the speaker’s credibility and good character, and (3) *pathos*, the evoking of the right emotions in the target audience. As Aristotle, we target argumentation that aims for persuasion — as opposed to critical discussions where *strategic maneuvering* is needed to achieve reasonableness (van Eemeren et al., 2014). Jovičić (2006) proposed a procedure to evaluate the effectiveness of persuasive argumentation, though not in a computational way.

In natural language processing, most computational argumentation research focuses on the mining of argumentative units and relations from text (Stab and Gurevych, 2014; Peldszus and Stede, 2015; Ajour et al., 2017). Argument mining infers the logical structure of arguments, but it does not analyze the strategy used to compose arguments. Feng and Hirst (2011) classify the five most common argumentation *schemes* of Walton et al. (2008). Such a scheme defines a pattern capturing the logical inference from an

argument's premises to its claim, i.e., it primarily aims at logos only. Song et al. (2014) combine the idea of schemes with strategy-related considerations, and Habernal and Gurevych (2015) annotate pathos in forum comments and blog posts. Likewise, Duthie et al. (2016) develop a corpus and an approach to mine explicit expressions of ethos from political debates, whereas Hidey et al. (2017) even annotate all three means of persuasion for the premises of arguments. The provided data and methods may be helpful for studying persuasive effectiveness, but they have not been employed so far for that purpose.

In (Wachsmuth et al., 2016), we use the output of argument mining to assess four quality dimensions of persuasive essays. Some assessed dimensions relate to rhetoric, such as argument strength, which is defined based on how many readers are persuaded. Habernal and Gurevych (2016) examine the reasons that make arguments convincing, and Lukin et al. (2017) study how effective logos-oriented and pathos-oriented arguments are depending on the target audience. However, none of these works considers the application of rhetorical strategies.

Tan et al. (2016) find that the chance to persuade someone in good-faith discussions on *Reddit Change My View* is increased through multiple interactions and an appropriate linguistic style. Cano-Basave and He (2016) analyze the impact of the semantic framing of arguments (e.g., “taking sides” and “manipulation”) in political debates. Similarly, Wang et al. (2017) reveal the importance of selecting the right framing of a discussion topic for winning classical debates. In such dialogical situations, the arguments of opposing parties are usually fragmented into several alternating parts. Our work, in contrast, analyzes the rhetorical strategies of complete monological texts where an author presents his or her entire argumentation.

Studying persuasive blog posts, Anand et al. (2011) develop a scheme with 16 persuasion tactics of four types: those that postulate outcomes of an uptake, those that generalize in some way, those that appeal to external authorities, and those that rely on interpersonal factors. These tactics are found in small text spans and could be seen as the local counterpart of the global strategies we consider. To our knowledge, only we have explicitly worked towards a computational analysis of such strategies so far. In particular, we presented a corpus with 300 news editorials whose units are labeled with their roles in the argumentation, such as “testimony” and “common ground” (Al-Khatib et al., 2016). In (Al-Khatib et al., 2017), we then trained a classifier on this corpus to find sequential role patterns in 30k New York Times editorials. While we observed insightful variances in the use of evidence across editorial topics, it still remains unclear to what extent such patterns really reflect rhetorical strategies.

Clearly diverging from previous research, we consider rhetorical strategies in argumentation *synthesis*, for which related work is generally still scarce. Bilu and Slonim (2016) generate new claims by recycling topics and predicates from existing claims, whereas Reisert et al. (2015) synthesize complete arguments based on the model of Toulmin (1958) where a claim is supported by data that is reasoned by a warrant. Like us, those authors rely on a pool of argument units from which arguments are built. However, they restrict their view to logical argument structure. The same holds for Green (2017) who generates arguments with particular schemes (e.g., “cause to effect”) that are used to teach learners how to argue.

Yanase et al. (2015) present a method that arranges the sentences of an argumentative text in a natural way. Sato et al. (2015) build upon this method. Their system pursues similar goals as we do, phrasing an ordered text with multiple arguments. We extend their idea by rhetorical considerations, and we propose a general argumentation synthesis model. It is in line with classical concepts of building natural language generation systems (Reiter and Dale, 1997), but it targets argumentative texts in particular. With that, we seek to contribute a missing aspect to technologies that synthesize arguments such as *The Debater* (Rinott et al., 2015), namely, how to deliver arguments *effectively*. So far, strategical systems exist only for formal argumentation (Rosenfeld and Kraus, 2016).

Our synthesis-oriented model covers content, structure, and style properties. For computational purposes, such properties need to be mined before. To capture content, some works identify the different frames under which a topic can be viewed (Naderi and Hirst, 2017), model key aspects of a topic (Menini et al., 2017), or analyze potentially strategy-related topic patterns in campaign speeches (Gautrais et al., 2017). In terms of structure, Persing et al. (2010) learn sequences of discourse functions in essays (e.g., “rebuttal” or “conclusion”) that correlate with a good organization, and we have modeled flows of arbitrary types of information to classify text properties, such as stance or quality (Wachsmuth and Stein, 2017).

In (Wachsmuth et al., 2017), we extend this idea to model sequential and hierarchical argumentation structure simultaneously. Regarding style, Lawrence et al. (2017) analyze rhetorical figures in arguments, Song et al. (2017) classify discourse modes in essays, and Zhang et al. (2017) study rhetorical questions in political discourse. All such methods may be relevant when we operationalize our model.

Finally, the use of strategies in synthesis scenarios follows up on early work on discourse planning (Young et al., 1994; Zukerman et al., 2000). The computational approaches relied on rule-based techniques to create effective arguments (Carenini and Moore, 2006) for a few selected tactics. More recent research on general text generation uses probabilistic models to employ text structure (Barzilay, 2010), or synthesizes texts such that they have a certain style in terms of sentiment or similar (Hu et al., 2017; Shen et al., 2017). Our model is meant to provide an abstract framework to be exploited in such approaches.

### 3 Model

We now delineate our model of rhetorical strategies for synthesizing a monological argumentative text following a specified rhetorical strategy. As clarified below, its three main building blocks follow Aristotle (2007). While we do not operationalize the model computationally in this paper, we will study its general adequacy with human experts in Section 5. An instance of the model has been given in Figure 1.

#### 3.1 Argumentation Synthesis following Rhetorical Strategies

Given a controversial topic or question (such as “Does college edu help financial security?” in Figure 1) and a stance towards it (such as “con”), the goal of a rhetorical strategy can be seen as delivering arguments along with facts, anecdotes, etc. to persuade some target audience of the intended stance in an effective way. In line with Aristotle (2007), we hypothesize that such a strategy can be manifested in a monological argumentative text through a purposeful encoding of the three means of persuasion (see Section 2). i.e., *logos*, *ethos*, and *pathos*. An instance of a strategy (which is tailored to a specific target audience) thus specifies to what degree each means should be present, such as *logos* 70%, *ethos* 10%, *pathos* 20%.<sup>1</sup>

For giving a persuasive speech, Aristotle (2007) postulates five consecutive *canons of rhetoric*: (1) *inventio*, the selection of arguments, (2) *dispositio*, the arrangement of the arguments to achieve maximum impact, (3) *elocutio*, the phrasing of the arguments in a clear and appropriate style, (4) *memoria*, the memorization of the speech, and (5) *actio*, the delivery of the speech with gestures, prosody, and further means. Since we focus on written argumentation, the latter two are obsolete in the given context. Accordingly, we model a rhetorical strategy for synthesizing monological argumentative texts as a process of the three steps mentioned in Section 1: selecting content, arranging structure, and phrasing style.

Technically, we regard a strategy as a script of operators from a set  $\Omega = \Omega_s \cup \Omega_a \cup \Omega_p$ , where  $\Omega_s$  are select operators,  $\Omega_a$  are arrange operators, and  $\Omega_p$  are phrase operators. These operators can be combined to form complex, nested  $n$ -ary operations on a set of input units. We expect that an applied strategy becomes manifest in measurable text properties at different granularity levels, ranging from single words and phrases over argumentative units and complete arguments, up to the full argumentation in a text. The identification of the best properties is beyond the scope of this paper, but the box in the lower right corner of Figure 1 summarizes an exemplary overall manifestation of a pathos and ethos-oriented strategy: The author focuses on the financial impact of college education to support her stance against it. She rebuts existing views, before she states her thesis and provides evidence, all in a personal evaluation.<sup>2</sup>

For a concrete operationalization of the model, the input to the operators may be pre-fabricated text units (roughly: clauses corresponding to single propositions), semantic representations in a suitable formalism (as in the case of classical “deep” text generation), or some intermediate form like partially-analyzed text units, where information about syntax, named entities, etc. is available. In line with the experiment reported in Section 5, we here assume that a set of candidate text units, which we call the *pool*, is available to make informed selections. Ideally, the units have been analyzed for (1) means of persuasion present, (2) stance on the topic, (3) framing of the topic, and (4) effectiveness when functioning in an argument.

<sup>1</sup>Also, a rhetorical strategy will have to consider conventions of the intended text genre (news editorial, persuasive essay, etc.), but we leave that aside for present purposes.

<sup>2</sup>The distinction of content, structure, and style may not be perfectly orthogonal in all cases, which is rooted in the complexity and polythetic nature of natural language. Notice in this regard that all labels depicted in Figure 1 are exemplary only.

### 3.2 Selecting the Content (Inventio)

The first step of synthesizing argumentation is to decide on those frames under which to view a topic that support the intended stance best, while matching the desired means of persuasion (as specified by the strategy). In accordance with these decisions, a set of units is to be selected from the pool. For instance, the editorial in Figure 1 frames college education in terms of its financial benefits and drawbacks, among other frames. A different strategy could put the focus on the freedom of career choice. When making the selection, facts, anecdotes, and similar statements may be added to the arguments, particularly to demonstrate credibility or to evoke emotions (if the strategy favors ethos or pathos, respectively). In Figure 1, we find a personal anecdote in the beginning, whereas the last paragraph gives statistical and testimonial evidence. Technically, the set  $\Omega_s$  will contain operators of two types:

$$frames : select(topic, stance, strategy) \quad units : select(frames, strategy)$$

To establish the knowledge base behind  $\Omega_s$ , a preceding corpus analysis will be needed, which associates frames and units for topic-stance combinations with information about their persuasive effectiveness and the encoded means of persuasion. As summarized in Section 2, several computational techniques to implement various steps of this analysis have been proposed, such as unit segmentation (Ajjour et al., 2017), frame classification (Naderi and Hirst, 2017), and key aspect clustering (Menini et al., 2017).

### 3.3 Arranging the Structure (Dispositio)

Given the units, the next step is to compose the selected units in structured arguments and to sequentially order them, in a way that maximizes their impact. The financial drawback argument in Figure 1, for instance, gives the debts from student loans as the premise for the author’s claimed solvency problems. The resulting argument (pro the intended stance) rebuts the preceding con argument and — probably for rhetorical reasons — is immediately followed by the main thesis within the sequential flow of the editorial (indicated by the vertical lines). To illustrate the means of persuasion, an emphasis on pathos could, for example, lead to a preference for gradually increasing the strength of emotional appeal throughout the text. In contrast, for a logos-oriented strategy, it may be important that the sequence of units *coheres* locally and globally (which for a pathos-oriented argument may be less relevant, or even detrimental). Technically, we distinguish two types of arrange operators in  $\Omega_a$ :

$$arguments : arrange(units, strategy) \quad flow : arrange(arguments, strategy)$$

For executing this step, knowledge about the roles of units in arguments (Stab and Gurevych, 2014) and about effective flows of text units (Wachsmuth and Stein, 2017) is needed. This script-like knowledge fuses the information on the means of persuasion and of unit strength on the one hand with principles of text coherence and rhetorical organization on the other.

### 3.4 Phrasing the Style (Elocutio)

Finally, the sequence of units is to be presented in an argumentative text, making stylistic decisions, which in turn are governed by the strategy. The technical approach depends much on the type of “unit” that is being implemented (see end of Section 3.1). For minimally-analyzed text units, one can add connectives indicating the relation between units. If deeper analyses are available, it may be possible to use markers encoding evidence types or discourse modes, among other rhetorical devices. If syntax can be controlled, then a pathos-oriented argument may prefer relatively flat structures, while a logos-oriented one may lean towards more complex embedding, inviting the reader to follow the path of reasoning. In the mentioned pro argument from the editorial, a concession is expressed through the word “while”, indicating that the con argument is outweighed by the pro argument. A personal anecdote is then given as a sarcastic (“thanks to”) causal evidence, which is used as a pathos-oriented persuasion attempt. Technically, we thus consider the set  $\Omega_p$  to be composed of at least two basic types:

$$sentences : phrase(units, arguments, strategy) \quad text : phrase(sentences, flow, strategy)$$

Existing techniques to identify the right phrasing operators include evidence type classification (Rinott et al., 2015; Al-Khatib et al., 2017), discourse mode identification (Song et al., 2017), and others.

Question (with Marked Controversial Topic)	# Texts
Should the fine for leaving <b>dog excrements</b> on sidewalks be increased?	9
Should <b>shopping malls</b> generally be allowed to open on holidays and Sundays?	8
Should public health insurance cover treatments in complementary and <b>alternative medicine</b> ?	8
Should Germany introduce the <b>death penalty</b> ?	8
Should only those viewers pay a <b>TV licence fee</b> who actually want to watch programs offered by public broadcasters?	7
Should there be a cap on <b>rent increases</b> for a change of tenant?	6
Should all universities in Germany charge <b>tuition fees</b> ?	6
Should the statutory <b>retirement age</b> remain at 63 years in the future?	6
Should the the <b>morning-after pill</b> be sold over the counter at the pharmacy?	6
Should <b>intelligence services</b> be regulated more tightly by parliament?	4

Table 1: The 10 questions from the Arg-Microtexts corpus from which we use ADUs for our dataset.

## 4 Data

This section describes the dataset in terms of a pool of argumentative discourse units that we built in order to study the adequacy of the proposed model for strategical argumentation synthesis.

### 4.1 Source Texts of the Argumentative Discourse Units

We decided to work with the English version of the Arg-Microtexts corpus (Peldszus and Stede, 2016), which was designed to provide crisp argumentation in a “pro and con” manner. The corpus contains 112 short argumentative texts that have been written in response to 18 questions on different controversial topics. The stance of each text towards the topic is labeled (*pro* or *con*). Every text consists of 3 to 10 manually segmented *argumentative discourse units (ADUs)* with a mean of about 5. ADUs can range from a single clause to (in principle) multiple sentences, which together contribute a single function to the argumentative structure. The structure is inspired by the model of Freeman (2011) and allows for linked, convergent, and serial support, as well as two kinds of attack (rebuttals and undercutters).

From the annotated structure, we can directly derive the *thesis* of each text (aka the main claim) and, for each other ADU, the stance that it takes towards this thesis (*pro* or *con*). In total, the Arg-Microtexts corpus contains 576 ADUs (112 theses, 339 pros, and 125 cons).

### 4.2 Decontextualization of the Argumentative Discourse Units

For our purposes, we need ADUs to be as independent as possible of the context in which they originally occurred. At this point, the fact that the corpus texts are relatively short and hence free of complex inter-relationships, becomes helpful. We semi-automatically preprocessed each unit in four steps:

1. *Removal of sentence information.* We removed all capitalizations of words that were due to sentence beginnings as well as all sentence delimiters.
2. *Removal of unit prefixes.* We removed all discourse markers and connectives at unit beginnings. The top 10 terms were “but” (43 times), “and” (39), “as” (18), “however” (15), “besides” (11), “for” (11), “also” (1), “that’s why” (10), and “yet” (10). Together, they made up 61% of all removed prefixes.
3. *Removal of unit infixes.* We removed the following markers within units: “also” (16 times), “however” (14), “therefore” (10), “hence” and “thus” (3 each), as well as “though” (2).
4. *Pronoun resolution.* Finally, we resolved all pronouns in the units whose reference was unclear or ambiguous when given only the question of the respective text. For example, we replaced “they” by “murderers and rapists” in “the have taken or destroyed another life” from a text on the death penalty.

### 4.3 A Pool of 200 Argumentative Discourse Units for 10 Controversial Topics

The goal of our study is to imitate the computational synthesis of monological argumentative texts, given a topic, a stance towards the topic, and a strategy specification. To ensure a reasonably large pool of ADUs for synthesis for each topic, we restrict our view to the 10 most frequent questions in the Arg-Microtexts corpus. The number of texts for each of them is listed in Table 1 (controversial topics are marked bold).

Type	#	Argumentative Discourse Unit	Stance
Thesis	t1	Germany should by no means introduce capital punishment	Con
	t2	Germany should not introduce capital punishment	
	t3	the death penalty is a legal means that as such is not practicable in Germany	
	t4	the state ought to prevent murder - not avenge it	
Con	c1	criminals should not be put in luxury prisons	Con
	c2	many people think that a murderer has already decided on the life or death of another person	
	c3	murderers and rapists have taken or destroyed another life	
	c4	proponents of the death penalty count on its deterring effect as well as and the ultimate elimination of any potential threat	
Pro	p1	a death would not be of any more use to those affected and their relatives than if the felon receives a long sentence	Con
	p2	a door must remain open for making amends	
	p3	capital punishment is not a solution	
	p4	courts are subject to human error	
	p5	despite the death penalty there are significantly more homicides in the US than in Germany	
	p6	every human, even those who have committed a despicable crime, can bring themselves to regret and change their opinion	
	p7	everyone must be given the chance to hone their conscience and possibly make amends for their deed	
	p8	it is a much graver punishment to be imprisoned forever and be tortured by one's own thoughts than to be killed quickly and easily by an injection	
	p9	it turns out time and again that innocent people are convicted and executed	
	p10	no one can claim the right to rule over the life or death of another human being	
	p11	no one may have the right to adjudicate upon the death of another human being	
	p12	we don't live in medieval times anymore	
<b>Topic</b>	Should Germany introduce the death penalty?		

Table 2: The candidate thesis, con, and pro units for one topic-stance pair from the dataset we provide.

To reduce the bias of the expert’s personal opinions in the study, we use only theses for one stance towards the respective topic. In particular, we kept only those with the majority stance on the topic, resulting in at least four theses for each of the 10 topics.<sup>3</sup> We kept all other ADUs, but we inverted the stance of an ADU in case it referred to one of the discarded theses, because, for instance an ADU that is against a discarded con thesis can be assumed to be in favor of a pro thesis on the same topic.

As a result, we had at least four con and 12 pro units for the four theses on each topic. Additional units were discarded randomly to end up with an equal number of units for all 10 topics. The resulting distribution of the unit types matches the desired distribution in the source corpus. We ordered the theses, con units, and pro units alphabetically (to avoid ordering bias). Table 2 exemplarily shows the list of 20 units for one topic. Each such list represents the pool of candidate units for one task in our experiment.<sup>4</sup>

## 5 Experiments

We now report on the experiment that we carried out with human experts in order to study whether the general process of selecting, arranging, and phrasing defined by our model from Section 3 is suitable for a strategical synthesis of argumentative texts.

### 5.1 Experimental Set-up for Manual Argumentation Synthesis following Rhetorical Strategies

We hypothesized that humans (1) agree when synthesizing arguments for the same strategy more than for different ones, and (2) synthesize differently depending on their strategy, especially in terms of selection.

**Participants** A diverse set of 26 qualified experts participated in the study (10 female, 16 male) with diverse demographic backgrounds (9 from North America, 11 from Europe, 6 from Asia). 16 came from groups related to computational linguistics, while the others were writing experts, acquired on *upwork.com*. 16 had a PhD or master, 7 a bachelor, and 3 a high school degree. No author of this paper participated.

**Strategies** To emphasize the effects of a strategical synthesis, we consider only two somehow principled rhetorical strategies, for which we gave the following short intuitions:

<sup>3</sup>In case of the question “Should shopping malls generally be allowed to open on holidays and Sundays?”, pro and con theses were balanced (4 each). We chose con, so we ended up with five pro-topic and five con-topic cases in total.

<sup>4</sup>We provide the output of all main steps of the described dataset construction at: <http://arguana.com/data>

Unit	Strategy	Selection		Arrangement	
		Majority	$\kappa$	Majority	$\kappa$
Thesis	Logos-oriented	59%	* 0.25	51%	0.16
	Pathos-oriented	51%	* 0.10	57%	0.23
	<b>Logos to pathos</b>	<b>43%</b>	<b>-0.02</b>	<b>53%</b>	<b>0.18</b>
Con	Logos-oriented	52%	0.13	45%	0.11
	Pathos-oriented	49%	0.10	54%	0.17
	<b>Logos to pathos</b>	<b>46%</b>	<b>0.08</b>	<b>48%</b>	<b>0.15</b>
3 pros	Logos-oriented	20%	* 0.14	31%	0.04
	Pathos-oriented	21%	* 0.11	31%	0.04
	<b>Logos to pathos</b>	<b>18%</b>	<b>0.04</b>	<b>30%</b>	<b>0.07</b>
Overall	Logos-oriented	59%	* 0.16	48%	0.10
	Pathos-oriented	52%	* 0.11	54%	0.15
	<b>Logos to pathos</b>	<b>46%</b>	<b>0.04</b>	<b>49%</b>	<b>0.13</b>

Table 3: Majority and mean Cohen’s  $\kappa$  agreement in selecting and arranging units of each type within the arguments of each principled strategy and across (*logos to pathos*).

- *Logos-oriented*. Argue based on logical reasoning, which means to make rational and logical conclusions towards the intended stance on the given topic.
- *Pathos-oriented*. Argue based on emotional reasoning, which means to appeal to the emotions of the reader regarding the topic within your arguments.<sup>5</sup>

**Tasks** All participants had to create one short argumentative text for each topic-stance pair, five logos-oriented and five pathos-oriented texts. Pair-strategy combinations varied across participants, but were balanced in overall terms. For each pair, the participants first had to *select* one thesis unit, one con unit, and three pro units from our pool that they thought could best form a persuasive argument following the given strategy.<sup>6</sup> Then, they had to *arrange* these five units in the most suitable way they saw. Finally, they should *phrase* a coherent text by adding discourse markers and connectives (example terms were given) as well as punctuation marks before and after each unit. We provided a spreadsheet that made this process as easy as possible. The participants were not trained in order to avoid biasing them towards our hypotheses.

## 5.2 Agreement and Differences in Selection, Arrangement, and Phrasing

The results of our experiment is a benchmark dataset with 13 manually synthesized texts for each combination of topic-stance pair and principled strategy. We provide it at <http://arguana.com/data>.

**Agreement** To make strategy differences explicit, Table 3 shows the participants’ agreement in selecting units and in arranging the three unit types for arguments *within* each of the two strategies and *across* them (“logos to pathos”). Majority agreement is given when over 50% of all participants decided equally about choosing a unit, and the mean pairwise Cohen’s  $\kappa$  agreement captures the average agreement of two participants. It is expected that the  $\kappa$  values are low, since various meaningful arguments can be synthesized for each topic from our pool. Accordingly, other measures such as Fleiss’  $\kappa$  are not suitable.<sup>7</sup>

For the *selection*, the decisive observation is that the  $\kappa$  is much higher within each strategy (e.g., 0.25 for logos and 0.10 for pathos in case of *thesis*) than across (-0.02).  $\kappa$  differences between each strategy and across marked with \* are significant with at least  $p < 0.05$  ( $t$ -test in case of normal distribution, Wilcoxon test otherwise). This speaks for the truth of Hypothesis 1, i.e., unit selection depends on the strategy. The majority agreement indicates that, especially for logos-oriented arguments, *thesis* (59%) and *con* (52%) are often selected uniformly. In contrast, the agreement for *arrangement* is similar within and across strategies, rather deviating between logos and pathos. We further explore this finding below.

<sup>5</sup>We told the participants that the two strategies may overlap. Ethos-oriented argumentation was not considered, because the demonstration of credibility seems hard, given that we required the participants to build arguments from our given pool of units.

<sup>6</sup>We use units as proxies for frames in the selection, as we do not have information on the exact frames covered by the units.

<sup>7</sup>High  $\kappa$  values would even be doubtful, because they would imply that the participants often selected *exactly the same* units, which would render the potential of our pool of argumentative discourse units questionable.

Connection	Phrasing		
	Majority	$\kappa$	Top Category
Before initial thesis	94%	0.35	None
Before initial con	40%	-0.22	Comparison
Before initial pro	92%	-0.21	None
Thesis → pro	46%	-0.29	Comparison
Thesis → con	41%	-0.32	None
Con → thesis	57%	-0.24	None
Con → pro	48%	-0.37	Comparison
Pro → thesis	69%	-0.26	Contingency
Pro → con	39%	-0.25	Comparison
Pro → pro	37%	-0.18	Expansion

Table 4: Majority and mean Cohen’s  $\kappa$  agreement, as well as the top category of discourse markers and connectives in phrasing for each connection between unit types.



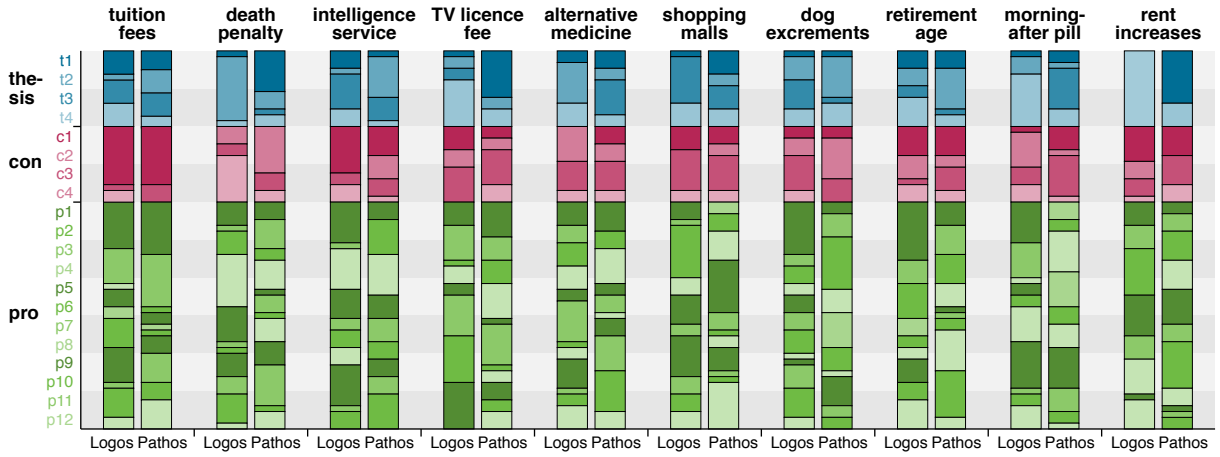


Figure 2: Distribution of the thesis units t1–t4 (blue), con units c1–c4 (red), and pro units p1–p12 (green) selected by all 13 annotators for each combination of topic and rhetorical strategy (logos/pathos-oriented).

Unit	Strategy	1	2	3	4	5	Mean	Unit Type Flow	Means	Frequency	Rank
Thesis	Logos	50%	21%	10%	2%	17%	2.1	(thesis, con, pro, pro, pro)	Logos	34.6%	1
	Pathos	56%	18%	5%	4%	17%	2.1		Pathos	43.1%	1
	<b>Overall</b>	<b>53%</b>	<b>19%</b>	<b>8%</b>	<b>3%</b>	<b>17%</b>	<b>2.1</b>		<b>Overall</b>	<b>38.8%</b>	<b>1</b>
Con	Logos	23%	42%	24%	8%	2%	2.3	(con, thesis, pro, pro, pro)	Logos	13.1%	2
	Pathos	25%	51%	17%	5%	2%	2.1		Pathos	13.8%	2
	<b>Overall</b>	<b>24%</b>	<b>47%</b>	<b>20%</b>	<b>7%</b>	<b>2%</b>	<b>2.2</b>		<b>Overall</b>	<b>13.5%</b>	<b>2</b>
Pro	Logos	9%	12%	22%	30%	27%	3.5	(thesis, pro, con, pro, pro)	Logos	12.3%	3
	Pathos	6%	11%	26%	30%	27%	3.6		Pathos	12.3%	3
	<b>Overall</b>	<b>8%</b>	<b>11%</b>	<b>24%</b>	<b>30%</b>	<b>27%</b>	<b>3.6</b>		<b>Overall</b>	<b>12.3%</b>	<b>3</b>

Table 5: Difference between strategies in the distribution of unit types over the five positions of the synthesized arguments, and mean rank of each type.

Table 6: Differences between strategies in the relative frequency and the rank of the three most common unit type flows in the synthesized arguments.

For *phrasing*, we just show in Table 4 to what extent the experts agreed in connecting consecutive unit types. We assume agreement if the same category of discourse marker or connective was used, distinguishing comparison, contingency, concession, expansion, other, and none. The majority agreement is high in many cases, while we see systematic  $\kappa$  disagreement. This implies that there were two opposing camps, although statistical reliability is limited. Quite intuitively, comparison terms (e.g., “although” or “but”) come often before or after con, whereas consecutive pros usually expand each other (e.g., with “and”).

**Differences** For brevity, we compare the relative frequencies of *selecting* each unit in logos-oriented and pathos-oriented arguments visually in Figure 2. For the theses, we observe clear strategy differences. For instance, nearly all logos texts on *death penalty* use t2, whereas the majority selected t1 for pathos there. In general, the thesis selection seems more uniform for logos. The distribution of con units is rather similar across strategies for many topics, suggesting that natural candidates to be rebutted exist among these. Conversely, the stacked bars of the pro units diverge often, as in the case of *dog excrements*. As far as our limited unit pool permits, these results support Hypothesis 2, i.e., different strategies lead to different selections of units and topic frames.

We analyze the *arrangement* in terms of the distribution of each unit type over the five positions in the synthesized texts as well as in terms of the resulting sequential unit type flows. As Table 5 reveals, the distributions are very stable across strategies. About half of all texts begin with the thesis, whereas pro units rather come at the end. Only for the con units, we see some differences, such as 42% (logos-oriented) versus 51% (pathos-oriented) at position 2. Accordingly, both strategies yield the same three most common flows, listed in Table 6, with (*thesis, con, pro, pro, pro*) at rank 1. We omit significance tests, since these results rather entail the conclusion that arrangement is not specific to the given strategies.

Strategy	Comparison		Contingency		Concession		Expansion		Other		None
	Top	Frequency	Top	Frequency	Top	Frequency	Top	Frequency	Top	Frequency	Frequency
Logos	but	20%	because	16%	indeed, of course	6%	and	15%	for one thing	2%	42%
Pathos	but	21%	because	18%	indeed, admittedly	8%	and	13%	for one thing	2%	39%
<b>Overall</b>	<b>but</b>	<b>20%</b>	<b>because</b>	<b>17%</b>	<b>indeed, admittedly</b>	<b>7%</b>	<b>and</b>	<b>14%</b>	<b>for one thing</b>	<b>2%</b>	<b>41%</b>

Table 7: The top terms and relative frequency of each discourse marker/connective category in phrasing.

Strategy	#	Text Manually Synthesized From Five Argumentative Discourse Units
Logos-oriented	t2	Germany should not introduce capital punishment.
	c4	Proponents of the death penalty count on its deterring effect as well as and the ultimate elimination of any potential threat.
	p5	<i>However</i> , despite the death penalty there are significantly more homicides in the US than in Germany.
	p4	<i>Furthermore</i> , courts are subject to human error.
	p9	It turns out time and again that innocent people are convicted and executed.
Pathos-oriented	t1	Germany should by no means introduce capital punishment.
	c2	Many people think that a murderer has already decided on the life or death of another person.
	p2	<i>Still</i> , a door must remain open for making amends.
	p6	Every human, even those who have committed a despicable crime, can bring themselves to regret and change their opinion.
	p11	<i>Therefore</i> , no one may have the right to adjudicate upon the death of another human being.

Table 8: Comparison of two selected argumentative texts, one for each strategy, manually synthesized based on the units from Table 2. The italicized discourse markers have been added by the participants.

Similar holds for the *phrasing* of argumentative texts, or at least for the given task of adding discourse markers and connectives. In particular, Table 7 presents the proportion of the most frequent categories of these. Logos-oriented and pathos-oriented arguments are practically identical in this regard, even the most often used terms in each category match. Comparisons (especially contrasts with “but”) as well as causal contingency markers dominate the applied phrasing operators.

**Insights** Exemplarily, Table 8 compares two arguments against the death penalty synthesized by different participants, one for each strategy. The theses differ in the emotional load rather than their framing. While the logos-oriented argument frames death penalty in terms of its potential deterrent effect and the execution of innocent people, the pathos-oriented puts full emphasis on the ability of humans to regret and change. The different numbers of frames lead to a slightly different phrasing with discourse markers. Matching our results, the arrangement of both arguments is the same on the abstraction level of unit types.

## 6 Conclusion

We propose a general model of synthesizing argumentation following rhetorical strategies in terms of Aristotle’s means of persuasion: logos, ethos, pathos. The model idealizes the synthesis as the selection, arrangement, and phrasing of argumentative discourse units (ADUs). Before we develop computational approaches based on the model, this paper has evaluated its general adequacy in an experiment with human experts. The results provide evidence that humans agree significantly more when synthesizing argumentative texts following the same strategy. In addition, we found that the arrangement of the ADUs and the re-phrasing of their connections is hardly affected by the strategy at all. A study of the phrasing of the actual units is left to future work. In the long term, we envisage a system that is able to automatically generate *effective* argumentation. Such a system requires two main types of resources: (1) a large pool of decontextualized ADUs covering diverse topics, and (2) a sufficiently flexible set of select, arrange, and phrase operators along with information about their effectiveness for specific topics and about the means of persuasion they encode. Given corpora with respective annotations, both resources can be developed using existing natural language processing techniques. We see this as the next step towards our goal.

**Acknowledgments** Thanks to Yamen Ajjour, Wei-Fan Chen, Yulia Clausen, Debopam Das, Erdan Genc, Tim Gollub, Yulia Grishina, Erik Hägert, Johannes Kiesel, Lukas Paschen, Martin Potthast, Robin Schäfer, Constanze Schmitt, Uladzimir Sidarenka, Shahbaz Syed, and Michael Völske for taking part in our study.

## References

- Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein. 2017. Unit segmentation of argumentative texts. In *Proceedings of the 4th Workshop on Argument Mining*, pages 118–128. Association for Computational Linguistics.
- Khalid Al-Khatib, Henning Wachsmuth, Johannes Kiesel, Matthias Hagen, and Benno Stein. 2016. A news editorial corpus for mining argumentation strategies. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3433–3443. The COLING 2016 Organizing Committee.
- Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein. 2017. Patterns of argumentation strategies across topics. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1351–1357. Association for Computational Linguistics.
- Jens Allwood. 2016. Argumentation, activity and culture. In *6th International Conference on Computational Models of Argument*, page 3.
- P. Anand, J. King, Jordan Boyd-Graber, E. Wagner, C. Martell, Douglas Oard, and Philip Resnik. 2011. Believe me — We can do this! Annotating persuasive acts in blog text. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.
- Aristotle. 2007. *On Rhetoric: A Theory of Civic Discourse* (George A. Kennedy, Translator). Clarendon Aristotle series. Oxford University Press.
- Regina Barzilay. 2010. *Probabilistic Approaches for Modeling Text Structure and Their Application to Text-to-Text Generation*, pages 1–12. Springer Berlin Heidelberg.
- Yonatan Bilu and Noam Slonim. 2016. Claim synthesis via predicate recycling. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 525–530. Association for Computational Linguistics.
- Amparo Elizabeth Cano-Basave and Yulan He. 2016. A study of the impact of persuasive argumentation in political debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1405–1413. Association for Computational Linguistics.
- Giuseppe Carenini and Johanna D. Moore. 2006. Generating and evaluating evaluative arguments. *Artificial Intelligence*, 170(11):925–952.
- Rory Duthie, Katarzyna Budzynska, and Chris Reed. 2016. Mining ethos in political debate. In *6th International Conference on Computational Models of Argument (COMMA 16)*, pages 299–310.
- 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*, pages 987–996. Association for Computational Linguistics.
- James B. Freeman. 2011. *Argument Structure: Representation and Theory*. Springer.
- Clément Gautrais, Peggy Cellier, René Quiniou, and Alexandre Termier. 2017. Topic signatures in political campaign speeches. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2342–2347. Association for Computational Linguistics.
- Nancy L. Green. 2017. Argumentation scheme-based argument generation to support feedback in educational argument modeling systems. *International Journal of Artificial Intelligence in Education*, 27(3):515–533.
- 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. Association for Computational Linguistics.
- Ivan Habernal and Iryna Gurevych. 2016. What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in web argumentation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1214–1223. Association for Computational Linguistics.
- Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*, pages 11–21.

- Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. 2017. Toward controlled generation of text. In *International Conference on Machine Learning*, pages 1587–1596.
- Ralph H. Johnson and J. Anthony Blair. 2006. *Logical Self-defense*. International Debate Education Association.
- Taeda Jovičić. 2006. The effectiveness of argumentative strategies. *Argumentation*, 20:29–58.
- John Lawrence, Jacky Visser, and Chris Reed. 2017. Harnessing rhetorical figures for argument mining. *Argument & Computation*, 8(3):289–310.
- Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument Strength is in the Eye of the Beholder: Audience Effects in Persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 742–753. Association for Computational Linguistics.
- Stefano Menini, Federico Nanni, Simone Paolo Ponzetto, and Sara Tonelli. 2017. Topic-based agreement and disagreement in us electoral manifestos. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2938–2944. Association for Computational Linguistics.
- Hugo Mercier and Dan Sperber. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34:57–111.
- Nona Naderi and Graeme Hirst. 2017. Classifying frames at the sentence level in news articles. In *Recent Advances in Natural Language Processing*, pages 536–542, 11.
- 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. Association for Computational Linguistics.
- Andreas Peldszus and Manfred Stede. 2016. An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: 1st European Conference on Argumentation (ECA 16)*. College Publications.
- 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. Association for Computational Linguistics.
- Paul Reisert, Naoya Inoue, Naoaki Okazaki, and Kentaro Inui. 2015. A computational approach for generating Toulmin model argumentation. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 45–55. Association for Computational Linguistics.
- Ehud Reiter and Robert Dale. 1997. Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87.
- Ruty Rinott, Lena Dankin, Carlos Alzate Perez, M. Mitesh 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*, pages 440–450. Association for Computational Linguistics.
- Ariel Rosenfeld and Sarit Kraus. 2016. Strategical argumentative agent for human persuasion. In *Proceedings of the 22nd European Conference on Artificial Intelligence*, pages 320–328.
- Misa Sato, Kohsuke Yanai, Toshinori Miyoshi, Toshihiko Yanase, Makoto Iwayama, Qinghua Sun, and Yoshiki Niwa. 2015. End-to-end argument generation system in debating. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pages 109–114. Association for Computational Linguistics and The Asian Federation of Natural Language Processing.
- Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In *Advances in Neural Information Processing Systems*, pages 6833–6844.
- Yi Song, Michael Heilman, Beata Beigman Klebanov, and Paul Deane. 2014. Applying argumentation schemes for essay scoring. In *Proceedings of the First Workshop on Argumentation Mining*, pages 69–78. Association for Computational Linguistics.
- Wei Song, Dong Wang, Ruiji Fu, Lizhen Liu, Ting Liu, and Guoping Hu. 2017. Discourse mode identification in essays. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 112–122. Association for Computational Linguistics.

- Christian Stab and Iryna Gurevych. 2014. Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 46–56. Association for Computational Linguistics.
- Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International World Wide Web Conference*, pages 613–624.
- Stephen E. Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.
- Frans H. van Eemeren, Bart Garssen, Erik C. W. Krabbe nad A. Francisca Snoeck Henkemans, Bart Verheij, and Jean H. M. Wagemans. 2014. *Handbook of Argumentation Theory*. Springer Netherlands.
- Henning Wachsmuth and Benno Stein. 2017. A universal model for discourse-level argumentation analysis. *ACM Transaction of Internet Technology*, 17(3):28:1–28:24.
- Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2016. Using argument mining to assess the argumentation quality of essays. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1680–1691. The COLING 2016 Organizing Committee.
- Henning Wachsmuth, Giovanni Da San Martino, Dora Kiesel, and Benno Stein. 2017. The impact of modeling overall argumentation with tree kernels. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2379–2389. Association for Computational Linguistics.
- Douglas Walton, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.
- Lu Wang, Nick Beauchamp, Sarah Shugars, and Kechen Qin. 2017. Winning on the merits: The joint effects of content and style on debate outcomes. *Transactions of the Association of Computational Linguistics*, 5:219–232.
- Toshihiko Yanase, Toshinori Miyoshi, Kohsuke Yanai, Misa Sato, Makoto Iwayama, Yoshiki Niwa, Paul Reisert, and Kentaro Inui. 2015. Learning sentence ordering for opinion generation of debate. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 94–103. Association for Computational Linguistics.
- R. Michael Young, Johanna D. Moore, and Martha E. Pollack. 1994. Towards a principled representation of discourse plans. *CoRR*, abs/cmp-lg/9406004.
- Justine Zhang, Arthur Spirling, and Cristian Danescu-Niculescu-Mizil. 2017. Asking too much? the rhetorical role of questions in political discourse. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1558–1572. Association for Computational Linguistics.
- Ingrid Zukerman, Richard McConachy, and Kevin B. Korb. 2000. Using argumentation strategies in automated argument generation. In *First International Conference on Natural Language Generation (INLG 00)*, pages 55–62.