Analyzing the Persuasive Effect of Style in News Editorial Argumentation

Roxanne El Baff$^{1,2}$ Henning Wachsmuth$^3$ Khalid Al-Khatib$^2$ Benno Stein$^2$

$^1$ German Aerospace Center (DLR), Germany, roxanne.elbaff@dlr.de
$^2$ Bauhaus-Universität Weimar, Weimar, Germany, <first>.<last>@uni-weimar.de
$^3$ Paderborn University, Paderborn, Germany, henningw@upb.de

Abstract

News editorials argue about political issues in order to challenge or reinforce the stance of readers with different ideologies. Previous research has investigated such persuasive effects for argumentative content. In contrast, this paper studies how important the style of news editorials is to achieve persuasion. To this end, we first compare content- and style-oriented classifiers on editorials from the liberal NYTimes with ideology-specific effect annotations. We find that conservative readers are resistant to NYTimes style, but on liberals, style even has more impact than content. Focusing on liberals, we then cluster the leads, bodies, and endings of editorials, in order to learn about writing style patterns of effective argumentation.

1 Introduction

The interaction between the author and the intended reader of an argumentative text is encoded in the linguistic choices of the author and their persuasive effect on the reader (Halmari and Virtanen, 2005). News editorials, in particular, aim to challenge or to reinforce the stance of readers towards controversial political issues, depending on the readers’ ideology (El Baff et al., 2018). To affect readers, they often start with an enticing lead paragraph and end their argument with a “punch” (Rich, 2015).

Existing research has studied the persuasive effect of argumentative content and structure (Zhang et al., 2016; Wachsmuth et al., 2016) or combinations of content and style (Wang et al., 2017; Persing and Ng, 2017). In addition, some works indicate that different types of content affect readers with different personalities (Lukin et al., 2017) and beliefs (Durmus and Cardie, 2018). However, it remains unexplored so far what stylistic choices in argumentation actually affect which readers. We expect such choices to be key to generating effective argumentation (Wachsmuth et al., 2018).

This paper analyzes the persuasive effect of style in news editorial argumentation on readers with different political ideologies (conservative vs. liberal). We model style with widely-used features capturing argumentativeness (Somasundaran et al., 2007), psychological meaning (Tausczik and Pennebaker, 2010), and similar (Section 3). Based on the NYTimes editorial corpus of El Baff et al. (2018) with ideology-specific effect annotations (Section 4), we compare style-oriented with content-oriented classifiers for persuasive effect (Section 5).\(^1\)

While the general performance of effect prediction seems somewhat limited on the corpus, our experiments yield important results: Conservative readers seem largely unaffected by the style of the (liberal) NYTimes, matching the intuition that content is what dominates opposing ideologies. On the other hand, the style features predict the persuasive effect on liberal readers even better than the content features — while being complementary. That is, style matters as soon as ideology matches.

Knowing about the specific structure of news editorials, we finally obtain common stylistic choices in their leads, bodies, and endings through clustering. From these, we derive writing style patterns that challenge or reinforce the stance of (liberal) readers of (liberal) news editorials, giving insights into what makes argumentation effective.

2 Related Work

Compared to other argumentative genres (Stede and Schneider, 2018), news editorials use many rhetorical means to achieve a persuasive effect on readers (van Dijk, 1995). Computational research has dealt with news editorials for retrieving opinions (Yu and Hatzivassiloglou, 2003; Bal, 2009), mining arguments (Al-Khatib et al., 2017), and

\(^{1}\)For reproducibility, the code of our experiments can be found here: https://github.com/webis-de/acl20-editors-style-persuasive-effect
analyzing their properties (Bal and Dizier, 2010; Scheffler and Stede, 2016). While Al-Khatib et al. (2016) modeled the structure underlying editorial argumentation, we use the corpus of El Baff et al. (2018) meant to study the persuasive effects of editorials depending on the readers’ political ideology. Halmari and Virtanen (2005) state that four aspects affect persuasion in editorials: linguistic choices, prior beliefs of readers, prior beliefs and behaviors of authors, and the effect of the text.

Persuasive effectiveness reflects the rhetorical quality of argumentation (Wachsmuth et al., 2017). To assess effectiveness, Zhang et al. (2016) modeled the flow of content in debates, and Wachsmuth et al. (2016) the argumentative structure of student essays. Others combined different features for these genres (Persing and Ng, 2015). The impact of content selection relates to the notion of *framing* (Ajjour et al., 2019) and is well-studied in theory (van Eemeren, 2015). As Wang et al. (2017), however, we hypothesize that content and style achieve persuasion jointly. We target argumentative style here primarily, and we analyze its impact on liberal and conservative readers.

In related work, Lukin et al. (2017) found that emotional and rational arguments affect people with different personalities, and Durmus and Cardie (2018) take into account the religious and political ideology of debate portal participants. In follow-up work, Longpre et al. (2019) observed that style is more important for decided listeners. Unlike them, we focus on the stylistic choices made in well-planned argumentative texts.

The lead paragraphs and the ending of an editorial have special importance (Rich, 2015). Hynds (1990) analyzes how leads and endings changed over time, whereas Moznette and Ruick (1968) examined the readability of an editorial based on them. To our knowledge, however, no one investigated their importance computationally so far. In this paper, we close this gap by analyzing what style of leads and endings is particularly effective compared to the editorial’s body.

3 Style Features

To model style, we need to abstract from the content of a news editorial. This section outlines the feature types that we employ for this purpose. Most of them have been widely used in the literature. Table 1 summarizes all features.

<table>
<thead>
<tr>
<th>Feature Base</th>
<th>Overview</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic inquiry and word count</td>
<td>Psychological meaningfulness in percentile</td>
<td>Pennebaker et al. (2015)</td>
</tr>
<tr>
<td>NRC emotional and sentiment lexicon</td>
<td>Count of emotions (e.g., sad, etc.) and polarity words</td>
<td>Mohammad and Turney (2013)</td>
</tr>
<tr>
<td>Webis Argumentative Discourse Units</td>
<td>Count of each evidence type (e.g., statistics)</td>
<td>Al-Khatib et al. (2017)</td>
</tr>
<tr>
<td>MPQA Arguing Lexicon</td>
<td>Count of 17 types of arguing (e.g., assessments)</td>
<td>Somasundaran et al. (2007)</td>
</tr>
<tr>
<td>MPQA Subjectivity Classifier</td>
<td>Count of subjective and objective sentences</td>
<td>Riloff and Wiebe (2003)</td>
</tr>
</tbody>
</table>

Table 1: Summary of the style feature types in our dataset. Each feature is quantified at the level of the editorial.

**LIWC** Psychological word usage is reflected in the Linguistic Inquiry and Word Count (Tausczik and Pennebaker, 2010). LIWC is a lexicon-based text analysis that assigns words to psychologically meaningful categories (Tausczik and Pennebaker, 2010). We use the LIWC version of Pennebaker et al. (2015), which contains 15 dimensions listed in the following with examples.

1. *Language metrics*: words per sentence, long words.
2. *Function words*: pronouns, auxiliaries.
3. *Other grammar*: common verbs, comparisons.
4. *Affect grammar*: positive and negative emotion.
5. *Social word*: family, friends.
8. *Biological processes*: body, health.
9. *Core drives and needs*: power, reward focus.
10. *Time orientation*.
12. *Personal concerns*.
13. *Informal speech*.
15. *Summary variables*.

The last dimension (15) contains four variables, each of which is derived from various LIWC dimensions:

(a) *Analytical thinking* (Pennebaker et al., 2014): The degree to which people use narrative language (low score), or more logical and formal language (high score).
(b) *Clout* (Kacewicz et al., 2014): The relative social status, confidence, and leadership displayed in a text.
(c) *Authenticity* (Newman et al., 2003): The degree to which people reveal themselves authentically.
(d) *Emotional tone* (Cohn et al., 2004): Negative emotions, for scores lower than 50, and positive emotions otherwise.

**NRC Emotion & Sentiment** To represent the mood of editorials, we use the NRC lexicon of Mohammad and Turney (2013). NRC contains a set of English words and their associations with (1) emotions such as *anger, disgust, and fear* as
well as (2) negative and positive sentiment polarities. These features are represented as the count of words associated with each category.

**Webis ADUs** To identify argumentative units in editorials that present evidence, we use the pre-trained evidence classifier of Al-Khatib et al. (2017). For each editorial, we identify the number of sentences that manifest anecdotal, statistical, and testimonial evidence respectively.

**MPQA Arguing** Somasundaran et al. (2007) constructed a lexicon that includes various patterns of arguing such as assessments, doubt, authority, emphasis. For each lexicon, we have one feature that represents the count of the respective pattern in an editorial.

**MPQA Subjectivity** We apply the subjectivity classifier provided in OpinionFinder 2.0 (Riloff and Wiebe, 2003; Wiebe and Riloff, 2005) on the editorials, in order to count the number of subjective and objective sentences there.

4 Data

As the basis of our analysis, we use the Webis-Editorial-Quality-18 corpus (El Baff et al., 2018). The corpus includes persuasive effect annotations of 1000 English news editorials from the liberal New York Times (NYTimes). The annotations capture whether a given editorial challenges the prior stance of readers (i.e., making them rethink it, but not necessarily change it), reinforces their stance (i.e., helping them argue better about the discussed topic), or is ineffective for them. Each editorial has been annotated by six annotators: three with liberal and three with conservative ideology.

To evaluate an editorial’s persuasive effect on liberals, we computed the majority vote of their annotations for the editorial (and, similarly, for conservatives). We ended up with 979 editorials with effect labels for liberals and conservatives. We found 21 duplicate editorials with the same content but different IDs (for these, we use the majority vote across all duplicates).

The corpus does not have predefined evaluation datasets. To mimic real-life scenarios, we chronologically split it into a training set (oldest 80%) and a test set (newest 20%). Table 2 shows the distribution of ideology-specific effects in the datasets.

<table>
<thead>
<tr>
<th>Class</th>
<th>Training Liberal</th>
<th>Training Conserv.</th>
<th>Test Liberal</th>
<th>Test Conserv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenging</td>
<td>126</td>
<td>128</td>
<td>22</td>
<td>41</td>
</tr>
<tr>
<td>Ineffective</td>
<td>118</td>
<td>292</td>
<td>32</td>
<td>71</td>
</tr>
<tr>
<td>Reinforcing</td>
<td>539</td>
<td>363</td>
<td>142</td>
<td>84</td>
</tr>
<tr>
<td>Overall</td>
<td>783</td>
<td>783</td>
<td>196</td>
<td>196</td>
</tr>
</tbody>
</table>

Table 2: Distribution of the majority persuasive effect of the news editorials in the given training and test set for liberal and conservative ideology respectively.

5 Prediction of Persuasive Effects

To assess the impact of news editorial style on readers, we employ our style-based features on the task of predicting an editorial’s persuasive effect: Given either of the two ideologies (liberal or conservative), predict for each editorial whether it is challenging, reinforcing, or ineffective.

We developed separate prediction models for the effect on liberals and conservatives, respectively. For each style feature type and for their combinations, we trained one SVM model with a linear kernel on the training set using scikit-learn (Pedregosa et al., 2011).

Given the dataset split mentioned above (training set 80%, test set 20%), we tuned the SVM’s cost hyperparameter using grid search with 5-fold cross-validation on the training set. Since the distribution of effect labels is highly skewed, we set the hyperparameter class_weight to “balanced”. We then trained the best model on the whole training set and evaluated it on the test set. For comparison, we also built models for standard content features (lemma 1- to 3-grams), and we consider the random baseline that picks an effect class by chance.

For both ideologies, Table 3 reports the macro and micro F1-scores for the style features, their best-performing combination, the content features, and the best combination of content and style.

We computed significance using Wilcoxon’s test to reveal differences between each two approaches among best style, content, best content+style, and baseline. We obtained the means of F1-scores used in the significance tests by conducting five-fold cross-validation on the test set, using the same SVM hyperparameters as above.

---

3Best style liberals: LIWC, MPQA Subjectivity. Best style conservatives: NRC Emotion&Sentiment, Webis ADUs
4Content-style liberals: LIWC, MPQA Arguing, MPQA Subjectivity, Content. Conservatives: MPQA Arguing, Content
5A non-parametric test was needed, because a normal distribution was not given.
We consider the last paragraph as the ending in all cases.

(Rich, 2015), we split each editorial into lead, body, and ending. For each part, we separately perform importance of specific discourse parts of editorials.

Given the Clustering Styles of Discourse Parts

In general, the results indicate that the persuasive effect seems hard to predict on the given corpus. Still, we observe that the style features play a notable role in predicting the effect of editorials on liberals. They achieve a significantly better macro F1-score of 0.43 when combined with content compared to 0.36 when using content alone, at p < 0.05. On the other hand, the F1-scores of content (macro 0.37, micro 0.38) and style (both 0.36) in predicting the effect on conservatives, are insignificantly different even from the baseline (0.33, 0.34).

These results suggest that style is important as soon as the ideology of a reader matches the one of the news portal (at least, this holds for liberal ideology), but not if it mismatches (here, conservative).

6 Identification of Style Patterns

Observing that the style of NYTimes editorials affects liberal readers, we seek to learn what patterns of writing style makes their argumentation effective. To this end, we (1) abstract each discourse part of an editorial (lead, body, ending) into a style label using cluster analysis and (2) identify sequential patterns of style labels that are specific to challenging, ineffective, and reinforcing editorials.

Clustering Styles of Discourse Parts Given the importance of specific discourse parts of editorials (Rich, 2015), we split each editorial into lead, body, and ending. For each part, we separately perform three steps on the training set of the given corpus: 6

6The corpus of Sandhaus (2008) contains lead and paragraph annotations. The lead spans either the first two paragraphs (994 editorials), the first three (5), or the first only (1). We consider the last paragraph as the ending in all cases.

Table 3: Test set micro and macro F1-scores of each feature type and their best combinations in classifying the persuasive effect on liberals and conservatives. * and † indicate significant differences at p < 0.05 against the Random baseline and Content respectively.

<table>
<thead>
<tr>
<th>Features</th>
<th>Liberals</th>
<th>Conservatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Macro</td>
<td>Micro</td>
</tr>
<tr>
<td>LIWC</td>
<td>0.31</td>
<td>0.40</td>
</tr>
<tr>
<td>NRC Emotion&amp;Sentiment</td>
<td>0.33</td>
<td>0.39</td>
</tr>
<tr>
<td>Webis ADUs</td>
<td>0.28</td>
<td>0.36</td>
</tr>
<tr>
<td>MPQA Arguing</td>
<td>0.33</td>
<td>0.41</td>
</tr>
<tr>
<td>MPQA Subjectivity</td>
<td>0.33</td>
<td>0.38</td>
</tr>
<tr>
<td><strong>Best Style</strong></td>
<td>*0.38</td>
<td>*0.49</td>
</tr>
<tr>
<td>Content</td>
<td>0.36</td>
<td>*0.49</td>
</tr>
<tr>
<td><strong>Best Content+Style</strong></td>
<td>*0.43</td>
<td>*0.54</td>
</tr>
<tr>
<td>Random baseline</td>
<td>0.23</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Table 4: Distribution of clusters over the leads, bodies, and endings of challenging, ineffective, and reinforcing editorials in the training set. The clusters are labeled by their most discriminating features (ordered). ▲, ◄, ▼, and ▼▼ denote relatively high, medium, and (very) low scores. The highest value in each row is marked bold.

-----|--------|--------|--------|--------|
Lead | ▲ tone, ▼ authenticity | 0.15  | 0.12  | 0.11  |
    | ▼ tone, ▲ authenticity | 0.11  | 0.13  | **0.14** |
    | ▼ tone, ▲ authenticity | **0.20** | 0.09  | 0.15  |
    | ▼ tone, ▼ authenticity, ▼# words | 0.11  | 0.11  | **0.14** |
    | ▶ tone, ▲ authenticity | 0.06  | **0.18** | 0.14  |
    | ▲ tone, ▼ authenticity | 0.13  | 0.14  | **0.15** |
    | ▼ tone, ▼ authenticity, ▼# words | **0.24** | 0.23  | 0.17  |
Body | ▲ tone, ▼ authenticity | 0.17  | **0.25** | 0.13  |
    | ▼ tone, ▼ authenticity, ▼ relativity | 0.09  | 0.05  | **0.10** |
    | ▼ tone, ▼ authenticity, ▼ relativity | **0.13** | 0.10  | 0.09  |
    | ▼ tone, ▼ authenticity, ▼ relativity | 0.15  | 0.10  | **0.17** |
    | ▼ tone, ▼ authenticity, ▼# words | 0.17  | 0.15  | **0.18** |
    | ▼ tone, ▼ authenticity, ▼ relativity | **0.11** | **0.11** | **0.16** |
    | ▼ tone, ▼ authenticity | 0.18  | 0.21  | 0.19  |
End. | ▲ tone, ▼ authenticity, ▼# words | 0.10  | **0.11** | 0.07  |
    | ▲ tone, ▼ authenticity, ▼# words | 0.24  | 0.25  | 0.25  |
    | ▲ tone, ▼ authenticity, ▼# words | 0.15  | 0.15  | 0.14  |
    | ▼ tone, ▼ authenticity, ▼# words | 0.06  | **0.08** | **0.09** |
    | ▼ tone, ▼ authenticity, ▼# words | **0.21** | 0.12  | 0.17  |
    | ▼ tone, ▼ authenticity, ▼# words | **0.06** | **0.08** | 0.06  |
    | ▼ tone, ▼ authenticity, ▼# words | **0.17** | **0.19** | **0.22** |

1. Extract the style features from Section 3.
2. Perform a cluster analysis on the style features using cosine k-means. k is determined with the elbow method on the inertia of the clusters.
3. Derive cluster labels from the most discriminating features across clusters: For each cluster, we determine those 2–3 values (e.g., “high tone, low authenticity”) whose combination suffices to significantly distinguish a cluster from others. With high to very low, we mean here a feature has significantly higher or lower scores compared to other clusters. 7

Table 4 shows the distribution of lead, body, and ending clusters over challenging, ineffective, and reinforcing editorials.

For each discourse part, the most discriminating feature is tone, followed by authenticity. The former combines positive (higher scores) and neg-

7For each feature (e.g., tone), we measured significance using Anova (in case of homogeneity and normality) or Kruskal (otherwise). In the case of p < 0.05, we conducted post-hoc analysis (independent t-test in case of normality, Mann-Whitney otherwise) with Bonferroni correction for each cluster pair, and we calculated the effect size r. Based on the effect size values, we deduced the labels of each cluster and the relative differences between them (high to very low).
Figure 1: Sequences of lead, body, and ending styles most specific to challenging, ineffective, and reinforcing news editorials. The triangles denote whether the given style attribute is high, medium, or (very) low. The ordering of attributes reflects their importance.

Comparison of Style Patterns

**Challenging editorial**
- Lead: Authentic, Relativity
- Body: Authenticity, Tone
- Ending: Tone

**Ineffective editorial**
- Lead: Authentic, Relativity
- Body: Authenticity, Tone
- Ending: Tone

**Reinforcing editorial**
- Lead: Authentic, Relativity
- Body: Authenticity, Tone
- Ending: Tone

Excerpt of the news editorial “Indonesia’s Avian Flu Holdout”, challenging to liberal annorators.

Indonesia sent a chill through the World Health Organization recently when it refused to supply any more samples of the avian flu virus that has killed scores of its people. The move, which seemed aimed at gaining access to vaccines at an affordable price, threatens the global effort to track the virus and develop vaccines. But Indonesia has raised a valid point that needs to be addressed: if a pandemic should strike, poor countries would be left without protection. (…)

In a typical flu season, the key strains emerge from Asia, while the vaccines are sold primarily in the West. This has not caused a ruckus because most developing countries consider influenza one of their lesser health threats. But with rising fears of an avian flu pandemic, the dynamic has changed.

Indonesia decided to act after a foreign company announced work on a vaccine that would be based on its samples. Indonesia stopped cooperating with the W.H.O. and started negotiations to send future samples to another vaccine maker in return for technology that would allow Indonesia to make its own vaccine. (…)

The W.H.O. needs to work much harder to encourage the transfer of vaccine production technology to countries, like Indonesia, that have the technical ability to use it. That will increase the supply of vaccine and presumably bring prices down. Even then, we fear, there still won’t be enough.

Figure 2: Example of a challenging editorial, along with the styles observed for its lead, body, and ending.

While these insights are naturally still vague to some extent and require more analysis in follow-up research, they show a first way of capturing the style of editorial argumentation.

7 Conclusion

This paper analyzes the importance of news editorials style in achieving persuasive effects on readers with different political ideologies. We find evidence that style has a significant influence on how a (liberal) editorial affects a (liberal) reader. Inspired by the theory of the high importance of the lead and ending in writing editorials (Rich, 2015), we also reveal common effective and ineffective style sequences (lead-body-ending) statistically.

Our findings help to understand how effective argumentation works in the political sphere of editorial argumentation — and how to generate such argumentation. In related work, El Baff et al. (2019) revealed the impact of style features on generating pathos- and logos-oriented short argumentative texts based on the rhetorical strategies discussed by Wachsmuth et al. (2018). With the findings of this paper, we go beyond, defining the basis of a style-dependent generation model for more sophisticated argumentation, as found in news editorials.
References


Roxanne El Baff, Henning Wachsmuth, Khalid Al-Khatib, Manfred Stede, and Benno Stein. 2019. Computational argumentation synthesis as a language modeling task. In 12th International Natural Language Generation Conference. ACL.


