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-editorials-style-persuasive-effect

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3154–3160
July 5 - 10, 2020.

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

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

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 Rarick (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.

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 words*: positive and negative emotion. (5) *Social word*: family, friends. (6) *Cognitive processes*: discrepancies, certainty. (7) *Perceptual processes*: feeling, seeing. (8) *Biological processes*: body, health. (9) *Core drives and needs*: power, reward focus. (10) *Time orientation*. (11) *Relativity*. (12) *Personal concerns*. (13) *Informal speech*. (14) *Punctuation*. (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 displaced 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).[2] 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, because 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.

[2]For copyright reasons, the corpus provides only annotations for IDs of editorials. The actual texts of these editorials come from the NYTimes Annotated Corpus (Sandhaus, 2008).

Class	**Training**		**Test**	
	Liberal	**Conserv.**	**Liberal**	**Conserv.**
Challenging	126	128	22	41
Ineffective	118	292	32	71
Reinforcing	539	363	142	84
Overall	783	783	196	196

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 F_1-scores for the style features, their best-performing combination,[3] the content features, and the best combination of content and style.[4]

We computed significance using Wilcoxon's test to reveal differences between each two approaches among *best style*, *content*, *best content+style*, and *baseline*.[5] We obtained the means of F_1-scores used in the significance tests by conducting five-fold cross-validation on the test set, using the same SVM hyperparameters as above.

[3]Best style liberals: *LIWC, MPQA Subjectivity*. Best style conservatives: *NRC Emotion&Sentiment, Webis ADUs*

[4]Content+style liberals: *LIWC, MPQA Arguing, MPQA Subjectivity, Content*. Conservatives: *MPQA Arguing, Content*

[5]A non-parametric test was needed, because a normal distribution was not given.

	Liberals		Conservatives	
Features	**Macro**	**Micro**	**Macro**	**Micro**
LIWC	0.31	0.40	0.25	0.26
NRC Emotion&Sentiment	0.33	0.39	0.28	0.29
Webis ADUs	0.28	0.36	0.31	0.31
MPQA Arguing	0.33	0.41	0.29	0.29
MPQA Subjectivity	0.33	0.38	0.26	0.28
Best Style	*0.38	*0.49	0.36	0.37
Content	0.36	*0.49	**0.37**	**0.38**
Best Content+Style	*†**0.43**	*†**0.54**	0.36	0.36
Random baseline	0.23	0.26	0.33	0.34

Table 3: Test set micro and macro F_1-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.

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 F_1-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 F_1-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]

[6]The 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.

Part	**Cluster**	**Chall.**	**Ineff.**	**Reinf.**
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, ▲relativity	0.17	**0.18**	0.15
	▶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**

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.

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-

[7]For 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*).

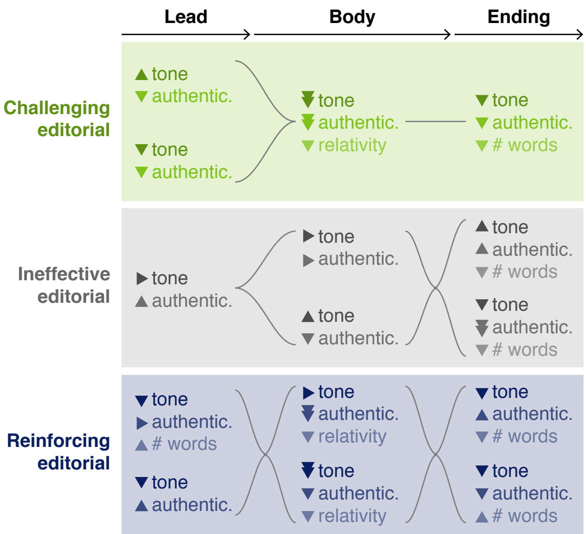

Figure 1: Sequences of lead, body, and ending styles most specific to challenging, ineffective, and reinforcing news editorials. The triangles denote whether the given style attribute is high, medium, or (very) low. The ordering of attributes reflects their importance.

ative (lower scores) emotional tones (Cohn et al., 2004). The latter indicates the degree to which people authentically reveal themselves; the higher the score, the more personal, humble, or vulnerable the writer is (Newman et al., 2003). In Table 4, we observe, for example, that the lead of challenging editorials over-proportionally often shows low authenticity, or that bodies with positive tone but low authenticity tend to be ineffective.

Identification of Style Patterns From Table 4, we determine the (maximum) two labels for each discourse part that are most specific to each of the three persuasive effect classes. From these, we build all possible lead-body-ending sequences, as visualized in Figure 1. According to a χ-square test, the distributions of these sequences differ significantly at $p < 0.05$. They reveal the following patterns of NYTimes editorials for liberal readers:

- *Challenging editorials* often begin with a polar emotional tone, followed by a negative tone. They tend to have low authenticity (i.e., not humble/personal) in the whole discourse (see Figure 2 for an example).
- *Ineffective editorials* over-proportionally often start with authenticity and dull tone. They then tend to diffuse in different directions and to have a short ending paragraph.
- *Reinforcing editorials* tend to start and end with a negative tone. They often avoid relativity in the actual arguments (i.e., in the body).

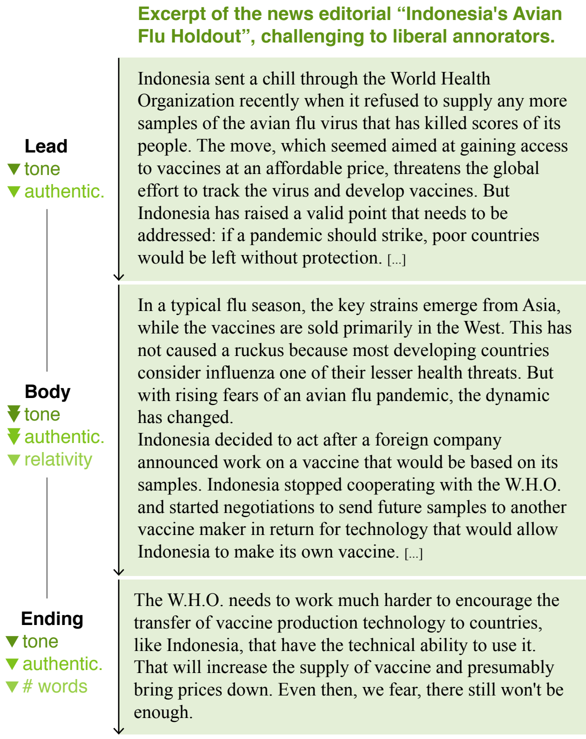

Figure 2: Example of a challenging editorial, along with the styles observed for its lead, body, and ending.

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

Yamen Ajjour, Milad Alshomary, Henning Wachsmuth, and Benno Stein. 2019. Modeling frames in argumentation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2922–2932, Hong Kong, China. Association for Computational Linguistics.

Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein. 2017. Patterns of argumentation strategies across topics. In *2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017*, pages 1362–1368. 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 *26th International Conference on Computational Linguistics (COLING 2016)*, pages 3433–3443. Association for Computational Linguistics.

Bal Krishna Bal. 2009. Towards an analysis of opinions in news editorials: How positive was the year? (project abstract). In *Proceedings of the Eight International Conference on Computational Semantics*, pages 260–263. Association for Computational Linguistics.

Bal Krishna Bal and Patrick Saint Dizier. 2010. Towards building annotated resources for analyzing opinions and argumentation in news editorials. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Languages Resources Association (ELRA).

Michael A Cohn, Matthias R Mehl, and James W Pennebaker. 2004. Linguistic markers of psychological change surrounding september 11, 2001. *Psychological science*, 15(10):687–693.

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.

Esin Durmus and Claire Cardie. 2018. Exploring the role of prior beliefs for argument persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1035–1045.

Frans H. van Eemeren. 2015. *Strategic Maneuvering*, pages 1–9. American Cancer Society.

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.

Roxanne El Baff, Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2018. Challenge or empower: Revisiting argumentation quality in a news editorial corpus. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 454–464. Association for Computational Linguistics.

Helena Halmari and Tuija Virtanen. 2005. *Persuasion across Genres: a Linguistic Approach*, volume 130. John Benjamins Publishing.

Ernest C Hynds. 1990. Changes in editorials: A study of three newspapers, 1955–1985. *Journalism Quarterly*, 67(2):302–312.

Ewa Kacewicz, James W Pennebaker, Matthew Davis, Moongee Jeon, and Arthur C Graesser. 2014. Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2):125–143.

Liane Longpre, Esin Durmus, and Claire Cardie. 2019. Persuasion of the undecided: Language vs. the listener. In *Proceedings of the 6th Workshop on Argument Mining*, pages 167–176, Florence, Italy. Association for Computational Linguistics.

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.

Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

James Moznette and Galen Rarick. 1968. Which are more readable: Editorials or news stories? *Journalism Quarterly*, 45(2):319–321.

Matthew L Newman, James W Pennebaker, Diane S Berry, and Jane M Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5):665–675.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical report, University of Texas at Austin.

James W Pennebaker, Cindy K Chung, Joey Frazee, Gary M Lavergne, and David I Beaver. 2014. When small words foretell academic success: The case of college admissions essays. *PloS one*, 9(12):e115844.

Isaac Persing and Vincent Ng. 2015. Modeling argument strength in student essays. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 543–552. Association for Computational Linguistics.

Isaac Persing and Vincent Ng. 2017. Lightly-supervised modeling of argument persuasiveness. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 594–604, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Carole Rich. 2015. *Writing and reporting news: A coaching method*. Cengage Learning.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*.

Evan Sandhaus. 2008. The New York Times Annotated Corpus. *Linguistic Data Consortium, Philadelphia*, 6(12):e26752.

Tatjana Scheffler and Manfred Stede. 2016. Realizing argumentative coherence relations in German: A contrastive study of newspaper editorials and Twitter posts. In *Proceedings of the COMMA Workshop: Foundations of the Language of Argumentation*, pages 73–80.

Swapna Somasundaran, Josef Ruppenhofer, and Janyce Wiebe. 2007. Detecting arguing and sentiment in meetings. In *Proceedings of the SIGdial Workshop on Discourse and Dialogue*, volume 6.

Manfred Stede and Jodi Schneider. 2018. *Argumentation Mining*. Number 40 in Synthesis Lectures on Human Language Technologies. Morgan & Claypool.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

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, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187. Association for Computational Linguistics.

Henning Wachsmuth, Manfred Stede, Roxanne El Baff, Khalid Al Khatib, Maria Skeppstedt, and Benno Stein. 2018. Argumentation synthesis following rhetorical strategies. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3753–3765. Association for Computational Linguistics.

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 for Computational Linguistics*, 5:219–232.

Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *International conference on intelligent text processing and computational linguistics*, pages 486–497. Springer.

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*, pages 129–136. Association for Computational Linguistics.

Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil. 2016. Conversational flow in Oxford-style debates. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 136–141, San Diego, California. Association for Computational Linguistics.