

Controversy and Sentiment in Online News

Yelena Mejova*, Amy X. Zhang†, Nicholas Diakopoulos†, Carlos Castillo*

*Qatar Computing Research Institute, †University of Maryland, ‡MIT CSAIL
ymejova@qf.org.qa, axz@mit.edu, nicholas.diakopoulos@gmail.com, chato@acm.org

采用数据驱动的方式，理解在新闻中，争议如何与如何与情感表达和偏见语言相互作用

ABSTRACT

How do news sources tackle controversial issues? In this work, we take a data-driven approach to understand how controversy interplays with emotional expression and biased language in the news. We begin by introducing a new dataset of controversial and non-controversial terms collected using crowdsourcing. Then, focusing on 15 major U.S. news outlets, we compare millions of articles discussing controversial and non-controversial issues over a span of 7 months. We find that in general, when it comes to controversial issues, the use of negative affect and biased language is prevalent, while the use of strong emotion is tempered. We also observe many differences across news sources. Using these findings, we show that we can indicate to what extent an issue is controversial, by comparing it with other issues in terms of how they are portrayed across different media.

在报导争议性事件时，记者需要更加注意措辞

1. INTRODUCTION

One of the most vital functions of the news media is to serve as a place to critically examine and present information about social, political, economic, and ideological issues of the day. Many of these issues are controversial, in the sense that they provoke arguments in which people express strong opposing opinions [10].

For this reason, journalists must often take special precaution and make careful language choices when they write about controversial issues. This can often manifest as different ways of using language, for instance news sources will often use a series of terms to signal that a controversy exists, such as “outcry,” “furor,” and “uproar” [6]. It has been theorized that journalists can also become susceptible to the ideologies, attitudes, and pressures of their organization [13], as well as unstated rules and norms [8]. These inputs can influence the particular language used to discuss controversial issues within a particular news source. The difference in framing could be subtle enough to be unnoticeable to the casual reader. However, using computational techniques in textual analysis, we can analyze large datasets of articles for consistent differences in the way different news organizations write about controversial issues.

Our contribution. In this work, we quantify the use of emotional and biased language when presenting controversial issues in the news. We begin by building a list of controversial and non-controversial terms in current news in the U.S. using crowdsourcing techniques. Then, we perform a large-scale analysis of millions of news articles from 15 U.S.-based news sources. We focus on the expression of sentiment using a series of lexical resources containing words conveying positive and negative emotions; this automatic analysis helps reduce the inherent subjectivity of traditional content analysis methods.

We demonstrate that controversial issues in news can be characterized by the use of fewer positive words and a greater presence of negative words. This finding is consistent across different media sources and confirmed with 4 different sentiment lexicons. Interestingly, we find that the use of highly emotional terms (as opposed to mild ones) is *less* likely in the context of controversial topics, suggesting a self-moderation on the part of the news sources.

Additionally, we perform an analysis based on a vocabulary of words signaling bias obtained from discussions in Wikipedia, and find that these bias terms also tend to occur more frequently in articles mentioning controversial topics, and can serve as a fairly accurate predictor of the level of controversy.

情绪和偏见词语

The next section outlines previous work related to controversy in news media. Next, we describe our dataset of online news (Section 3) and describe the process we used to label strongly controversial, somewhat controversial, and non-controversial words (Section 4). We then compare controversial and non-controversial articles in terms of a series of bias and sentiment lexicons in Section 5, and discuss the differences in the strength with which annotators perceive a topic as controversial and the treatment it received in news media in Section 6. Lastly we discuss the implications and limitations of such computational approaches to media analysis.

2. PREVIOUS WORK

Controversy has been examined in both social media, including Wikipedia and Twitter, and more traditional news sites.

The unique structure of Wikipedia as a collaborative endeavor has been used by Rad and Barbosa [11] who detect controversy based on mutual reverts, bi-polarity in the collaboration network, and/or mutually-reinforced scores for editors and articles (controversial editors work in controversial articles). The fact that an article on Wikipedia is controversial has then been used to evaluate the level of controversy of other documents (e.g. web pages) by mapping them to related Wikipedia articles [9].

Taking a content-driven approach, Pennacchiotti and Popescu [10] detect controversies around celebrities in Twitter. They use a number of features including the presence of sentiment-bearing

words, swear words, and words in a list of controversial topics from Wikipedia.

需要调研

Controversies in online news. In news sites, Choi et al. [5] detect controversial issues by looking at words that frequently appear in contexts containing positive and/or negative sentiment words. Also using lexicons, Chimmalgi [4] study controversy in user comments of news articles. Taking a more data-driven approach, Awadallah et al. [1] describe a method in which opinion holders and their opinions in news articles are identified by an iterative method based on a seed set of patterns that describe expressions of support or opposition to an idea.

In contrast to previous work that applies sentiment analysis to controversies, we do not assume that sentiments and controversy are related, but instead demonstrate it experimentally, thus not only providing an empirical basis for future use of sentiment lexicons, but also discovering new insight into potential self-moderation of the news sources.

3. NEWS ARTICLES

Data was provided by NewsCred,¹ which aggregates news content from thousands of news sources, and makes it available via an API (described further in Diakopoulos et al. [7]). We chose 15 high-volume sources from NewsCred and considered their articles in the period from March to September 2013. The main criterion we used for this selection was variety—including both national and regional sources—while at the same time we attempted to keep the list relatively brief for ease of analysis and exposition: CNN, Reuters, Usa Today, Los Angeles Times, Washington Post, Chicago Tribune, News Day, Minneapolis Star Tribune, Houston Chronicle, Philadelphia Inquirer, Honolulu Star-Advertiser, Huffington Post, New York Times, ProPublica, and Talking Points Memo.

Two data gathering tasks were done. First, we collected a random subset of articles for the purposes of obtaining high-frequency words. This is the list of initial topics used for the labeling task described in Section 4. Second, we searched the selected media sources for articles containing each topic word using the NewsCred API. In total, over 21 million articles were collected, with an average of around 3,000 and median of 1,000 per topic per news source. Reuters was the most prolific source with an average of 11,659 articles per topic, and ProPublica the least at 31 articles.

4. CONTROVERSIAL WORDS

Whether a word is controversial or not is a highly subjective and context-dependent matter. Since controversy is socially constructed, we performed an annotation effort with a relatively large pool of annotators. We employed 25 annotators based in the U.S., hired through crowdsourcing platform CrowdFlower (<https://crowdflower.com/>) and paid \$0.80 (USD) for every 100 words annotated, following standard pricing practices of this platform.²

Initial Words. The initial list was composed of the top frequent words in a large random sample of articles. We filtered this list to keep only single-word nouns using a part-of-speech tagger. Next, we removed generic English stopwords, as well as frequent news-specific stopwords (*ap, broadcast, press, published, rewritten, redistributed, rights, copyright, reserved*), and kept the top 2,000 most frequent terms.

¹<http://newscred.com/>

²In total 40 annotators participated in the task, but only 25 of them contributed more than 100 labels. This is common in crowdsourcing platforms.

Table 1: List of words identified during the crowdsourcing task.

Strongly Controversial (145): *abuse, administration, afghanistan, aid, america, american, army, attack, attacks, authorities, authority, ban, banks, benefits, bill, bills, border, budget, campaign, candidate, candidates, catholic, china, chinese, church, concerns, congress, conservative, control, country, court, crime, criminal, crisis, cuts, debate, debt, defense, deficit, democrats, disease, dollar, drug, drugs, economy, education, egypt, election, elections, enforcement, fighting, finance, fiscal, force, funding, gas, government, gun, health, immigration, inaccuracies, india, insurance, investigation, investigators, iran, israel, job, jobs, judge, justice, killing, korea, labor, land, law, lawmakers, laws, lawsuit, leadership, legislation, marriage, media, mexico, military, money, murder, nation, nations, news, obama, offensive, officials, oil, parties, peace, police, policies, policy, politics, poll, power, president, prices, primary, prison, progress, race, reform, republican, republicans, restrictions, rule, rules, ruling, russia, russian, school, security, senate, sex, shooting, society, spending, strategy, strike, support, syria, syrian, tax, taxes, threat, trial, unemployment, union, usa, victim, victims, violence, vote, voters, war, washington, weapons, world*

Somewhat Controversial (45): *account, advantage, amount, attorney, chairman, charge, charges, cities, class, comment, companies, cost, credit, delays, effect, expectations, families, family, february, germany, goal, housing, information, investment, markets, numbers, oklahoma, parents, patients, population, price, projects, raise, rate, reason, sales, schools, sector, shot, source, sources, status, stock, store, worth*

Non-Controversial (272): *60s, 70s, addition, address, afternoon, agreed, amp, angeles, answer, april, attention, avenue, average, ball, base, bay, beach, beginning, bit, block, blue, bowl, box, boy, boys, brother, building, bus, call, calling, calls, camp, car, cars, central, cents, click, close, cloudy, club, coast, cup, dallas, date, daughter, davis, day, decade, decades, december, def, delivery, door, download, drive, eagles, end, entire, era, evening, face, faces, facility, fall, fans, father, feel, feeling, feet, fell, field, finish, floor, form, fort, francisco, friday, friend, friends, fun, girl, girls, ground, gt, guy, guys, half, hall, hand, hands, hawaii, heart, heat, heavy, hill, hits, hold, hopes, host, hotel, hour, hours, house, houston, hundreds, husband, ice, illinois, index, indiana, innings, island, january, johnson, jones, june, kansas, kind, lack, lake, leave, lee, letter, levels, light, line, lines, lot, lows, lt, main, make, makes, mark, mass, material, matter, medium, men, mid, middle, miles, mind, minneapolis, minutes, moment, monday, month, months, morning, mother, mountain, move, mph, museum, names, natural, net, night, north, note, notes, november, october, opening, park, part, parts, pass, period, person, philadelphia, pick, pitch, plant, play, player, playing, pm, point, post, practice, put, quarter, rain, read, reading, red, rest, restaurant, rise, rock, rose, round, sale, san, saturday, scene, search, season, seasons, seconds, selling, september, series, set, showers, showing, shows, sign, signs, smith, son, sox, special, spot, spring, square, stadium, stage, start, starting, starts, station, stay, step, stores, street, student, summer, sun, sunday, thing, things, thinking, thought, thousands, thunderstorms, thursday, time, title, top, total, transportation, type, unit, valley, vehicle, version, village, visit, wait, walk, wall, watch, water, ways, wednesday, week, weekend, weeks, williams, wind, winds, winner, winter, word, writer, yards, year, years, york*

Crowdsourced annotation. We used a four-point scale to classify 2,000 high-frequency terms, asking 7 “trusted” annotators per word the following question:

You need to be familiar with U.S. news media. Indicate if *word* is:

- (C3) Strongly Controversial: people often disagree and debate with opposing viewpoints.
- (C2) Somewhat Controversial: people sometimes disagree or have debates with opposing viewpoints.
- (C1) Less Controversial: people infrequently disagree or have debates with opposing viewpoints.
- (C0) Non-Controversial: people almost never disagree or debate with opposing viewpoints.

“Trusted” annotators are selected by including a set of terms for which the label was known. These terms were obtained by a preliminary task in which we asked 5 crowdsourcing workers for each of 1,000 words whether they believed it to be controversial in U.S. news media or not. We selected a balanced set of 94 words for which there was perfect agreement in the preliminary task and assigned them to C0 and C1 (for non-controversial terms) and C2 and C3 (for controversial terms). Annotators who did not agree substantially with this gold standard were not considered “trusted.”

Finally, we considered only labels for which the majority label was larger than 60% among the 7 workers. This yields Table 1 containing 145 controversial terms having label C3, 45 terms having label C2, and 272 terms having label C0. Class C1 is made of

borderline cases and elicited very little agreement, it was thus not considered.

Controversial terms in Table 1 are part political terms referring to *congress* and *legislation*, social terms like *education* and *unemployment*, country names including *russia* and *china*, and terms such as *criminal* and *threat*. Those in the medium category may potentially belong to a controversial topic, such as *february* being Black History Month in the U.S. Finally, the non-controversial terms include mostly generic words.

5. CONTROVERSY AND SENTIMENT

Using the dataset we describe above, we explore the vocabulary used in articles mentioning controversial/non-controversial topics using a series of lexical resources related to sentiment and bias.

Method. The general methodology is to consider in turn each word w in our vocabulary of 462 topics, and each media source s in our list of 15 news media. Then, collect all articles in the source s containing the word w , eliminate duplicate and near-duplicate articles (typically a by-product of articles that have multiple URLs), and combine all the articles in one topical (and very large) super-article for analysis.

The analysis consists of examining this content using a lexical resource, counting the proportion of words matching a certain category in the lexical resource. We consider the following sentiment lexicons:

- Affective Norms for English Words³ (ANEW) is a set of normative emotional ratings for 2,476 English words. We use the “valence” rating considering positive (respectively, negative) the ratings above (respectively, below) the mean.
- General Inquirer⁴ is a list of 1,915 words classified as positive, and 2,291 words classified as negative.
- MicroWNOp⁵ [3] is a list of 1,105 WordNet *synsets* (cognitive synonyms) classified as positive, negative, or neutral.
- SentiWordNet⁶ [2] assigns to each synset of WordNet (around 117,000) a positive and negative score determined by a diffusion process.

Additionally, we use a bias-specific lexicon:

- Bias Lexicon⁷ is a list of 654 bias-related lemmas extracted from the edit history of Wikipedia by Recasens et al. [12]. Sentiment words are used as contributing features in the construction of this bias lexicon.

To assess how prominent bias- and sentiment-laden terms are in controversial topics, we count the number of times each lexicon term is used in each super-article, and divide this count by the total length of the article, resulting in a proportion of the text which uses the lexicon terms.

Bias, positive and negative emotions. The distributions of the proportion of words matching each vocabulary are shown in Figure 1 in which for brevity we have selected three media: Huffington Post (HUF), CNN, and Reuters (REU). As we show in the next section, these three news media sources differ in the extent to which the use of lexicon words are used around controversial words in their articles (from most to least).

First, we observe that *the use of bias terms is more likely in controversial topics than non-controversial topics*. This is statistically significant at $p < 0.01$ for all 15 news sources.

Second, *the use of negative terms is more likely in controversial topics than non-controversial topics*. This is statistically significant at $p < 0.01$ for 46 out of the 60 combinations of source and lexicon (15 sources and 4 sentiment lexicons), with 9 ties (no difference significant at this level in either direction) and 5 cases in which the difference is significant at this level in the opposite direction.

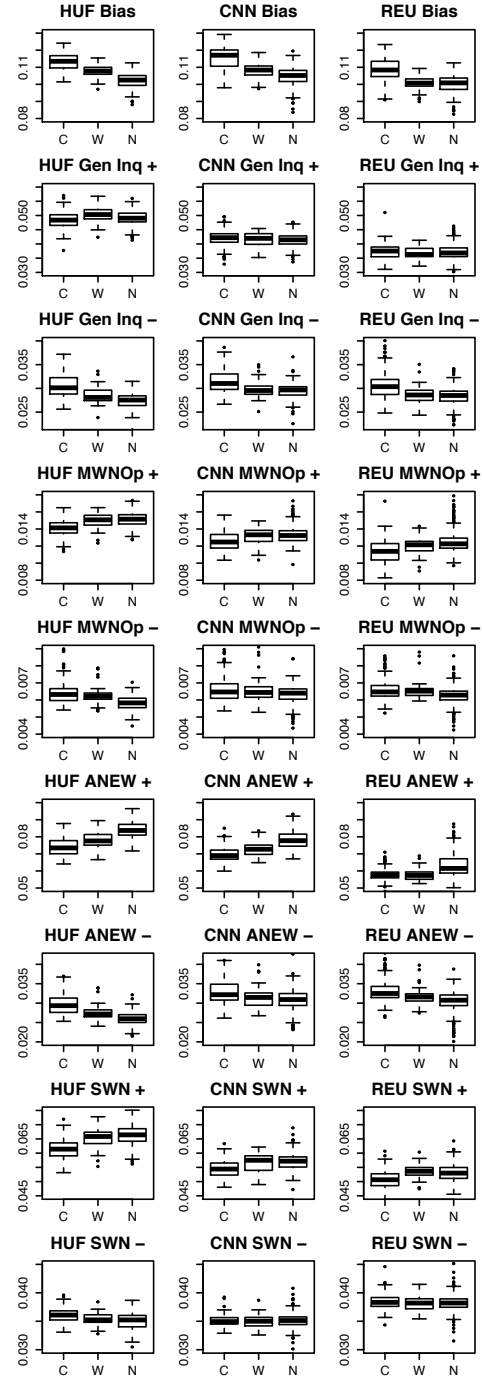


Figure 1: Distributions of the proportion of biased, positive (+), and negative (−) words in the Huffington Post (HUF), CNN, and Reuters (REU), across controversial topics (C), somewhat controversial topics (W) and non-controversial topics (N).

³<http://csea.phhp.ufl.edu/media/anevmessag.html>

⁴<http://www.wjh.harvard.edu/~inquirer/homecat.htm>

⁵<http://www-3.unipv.it/wnop/>

⁶<http://sentiwordnet.isti.cnr.it/>

⁷http://www.mpi-sws.org/~cristian/Biased_language.html

Table 2: Top news sources in terms of the difference in the use of bias/emotional words in controversial and non-controversial topics.

	Bias	Gen Inq <i>Strong</i>	ANEW –
1.	Huffington Post	Huffington Post	Huffington Post
2.	Washington Post	Washington Post	USA Today
3.	New York Times	New York Times	New York Times
4.	LA Times	CNN	Washington Post
5.	USA Today	LA Times	LA Times

Third, the use of positive terms is more likely in non-controversial topics than controversial topics. This is statistically significant at $p < 0.01$ for 40 out of the 60 conditions, with 16 ties and 4 cases with a significant difference in the opposite direction.

Strong emotions. Three of our lexicons (ANEW, MicroWNOp and SentiWordNet) include scores that allow us to distinguish between weakly and strongly emotional terms. We observe that the use of strong emotional words is less likely in controversial topics. This is statistically significant at $p < 0.01$ for 32 out of 45 conditions (15 sources and 3 sentiment lexicons), with only 1 of the remaining conditions having a significant difference in the other direction.

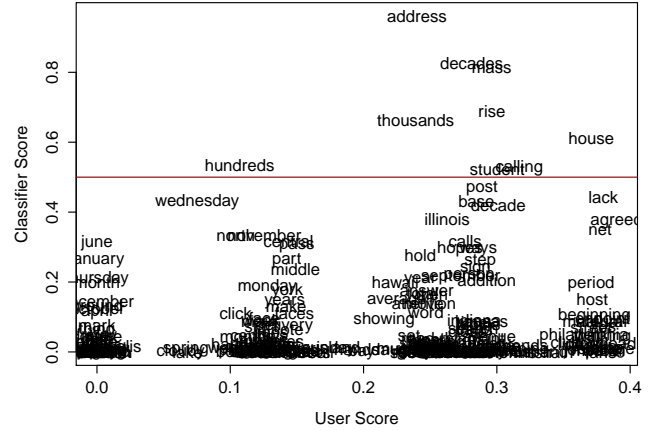
Differences across sources. We find a great variety in the different treatment that controversial and non-controversial topics have, in terms of the use of biased and emotional words. In terms of statistical significance, the clearer difference between controversial and non-controversial topics was observed using (i) the lexicon of bias words, (ii) the General Inquirer strong sentiment words, and (iii) the ANEW negative words.

We next rank the media sources in terms of their different usage of words in these lexicons in controversial and non-controversial topics. The top 5 sources for each one are shown on Table 2. We note that several media sources repeat in this list, with Huffington Post, Washington Post, and New York Times remaining on top, indicating a consistently large difference in their usage of sentiment words around controversial topics, compared to non-controversial ones.

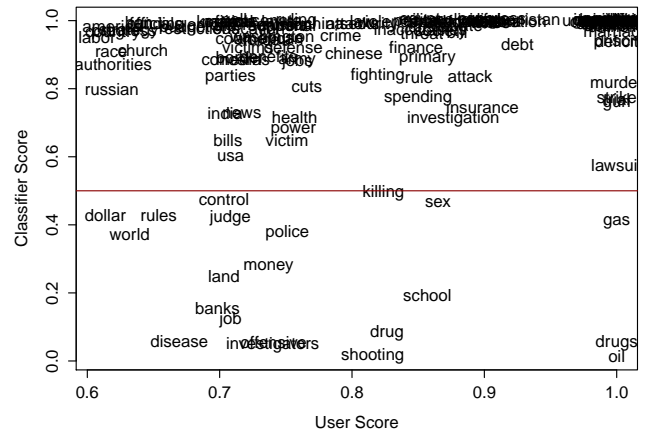
6. RANKING CONTROVERSY WORDS

Finally, we assign a score to each topic in our list of controversial and non-controversial terms by using logistic regression, using as input features the proportion of words from each lexicon in each news source, and as training data the manually-labeled words (using only the classes C3 and C0 that represent the extreme values). This is done using feature selection, selecting 5 features out of the total 195, and training a logistic regression classifier. Then, the same classifier is applied to the training data (the purpose of this is not to generalize to unseen topics, but to understand the existing one), and a score between 0 and 1 is computed for each word (0.0 is non-controversial, 0.5 is undecided, and 1.0 is strongly controversial).

Figure 2 depicts the training errors, which appear above the horizontal line in the plot for non-controversial topics and below it in the plot for controversial topics. To put these errors into context, we include in the figure the confidence of the manual annotation process. Note that most of the terms are classified correctly, appearing at the bottom of non-controversial and top of controversial figures. In the next section, we discuss the possible reasons for the misclassification of some of the topics.



(a) Non-controversial words; correctly classified words appear below the horizontal line.



(b) Controversial words; correctly classified words appear above the horizontal line.

Figure 2: Scores of controversial and non-controversial words including classification errors. “User score” is the confidence with which the manual labeling was done (with at least 7 annotators per element), while “classifier score” is the output of the classifier on the training data.

7. DISCUSSION

In this study, we found that controversial issues are often framed using negative emotions. Many of the terms that were found to be controversial are oftentimes related to social problems and violence, such as *gun*, *fighting*, *crime*, *victim* and *shooting*, or war, such as *strike*, *weapons*, and *army*. There may not be as many opportunities to discuss some of these issues in a non-negative light. Other terms could be viewed in a negative, positive, or even a neutral light depending on the context at hand, for instance, terms such as *health*, *security*, or *jobs*. However, because these words often invoke larger ideological issues of government spending, privacy, or the economy, it may be that frames emerge that are more antagonistic. For instance, instead of emphasizing the positive aspects of one’s view, a writer may choose to highlight the negative aspects of the opposing view.

Additionally, we found that controversial topics also involve less strongly emotional words. Theories of framing posit that various organizational pressures can shape the frames different news sources employ [13]. This may point to an effort on the part of news

agencies to more tightly control the language used when discussing controversial topics. As an overt example of organization influence on language use, many organizations today adhere to different well-known style guides when choosing certain language around sensitive topics. The use of different style guides may point to some of the similarities and differences we saw between sources and certain controversial words. Finally, the use of biased terms become more prevalent around controversial words, with abstract notions of fairness like *justice* and *rights*, and judgment-laden terms like *terrorist* and *criminal* having a high correlation.

Based on the above observations, we built a model to estimate to what degree a news source is treating a term as controversial. We find that some strongly controversial terms are instead classified as non-controversial—they appear in the bottom right corner of Figure 2(b). In some cases, terms that have different connotations depending on the context may have lead to incorrect classifications. For instance, *oil*, when referenced in a financial article may treat the subject objectively as opposed to using biased or emotional language. Another example is *drug* which can sometimes simply reference drug stores. However, the classification of other less ambiguous terms, such as *sex* and *killing* as non-controversial prompts re-examination of these topics as controversial. Further work is needed to model more clearly the topics represented by these terms in order to better understand their context.

The approach described in this work also allows us to examine the language of each agency around a controversy using a controlled vocabulary. For example, for the term *democrats* we compare the top 30 terms of the bias words lexicon used by each agency. We see that, along with more topical terms like *obama* and *democratic*, Huffington Post also includes more general and subjective terms like *very* and *good*, unlike, say, CNN and Reuters. The lexicons also allow us to glimpse a more local approach to news of smaller news agencies compared to national ones. When discussing *murder*, Reuters and CNN often mention a larger framework of *government* and other *groups*, whereas Philadelphia Inquirer, Honolulu Star-Advertiser, and Houston Chronicle mention particular people (*woman*, *victim*) and places (*university*, *west*). Further development of lexical resources for deeper understanding of news coverage beyond sentiment is an exciting future direction of this research.

The large-scale analysis we have conducted is an initial inquiry into quantifying how news sources differ in their framing. By being able to pick apart how news sources differ in often subtle ways, we can begin to uncover implicit biases in framing or language use by different news entities. That many of these news sources are quite large, employing hundreds and thousands of writers and editors, suggests that organization-level pressures to conform to a particular standard or world view may in fact exist. This work can also be used to inform automatic style guide checkers, serve as a reference to journalists interested in maintaining objectivity, or readers wishing to monitor their news intake.

Data release: our list of strongly controversial, somewhat controversial and non-controversial terms, along with their scores is available at www.yelenamejova.com/resources/controversial_words.txt.

Acknowledgments: the authors would like to thank NewsCred for making the data available for our research.

References

- [1] R. Awadallah, M. Ramanath, and G. Weikum. Harmony and dissonance: Organizing the people’s voices on political controversies. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM ’12, pages 523–532, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-0747-5.
- [2] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204, 2010.
- [3] S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, and G. Gandini. *Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining*. 2007.
- [4] R. V. Chimmalgi. Controversy trend detection in social media. Master’s thesis, Louisiana State University, May 2010. rajshekhar_2010_controversy.
- [5] Y. Choi, Y. Jung, and S.-H. Myaeng. Identifying controversial issues and their sub-topics in news articles. In H. Chen, M. Chau, S.-h. Li, S. Urs, S. Srinivasa, and Wang, editors, *Intelligence and Security Informatics*, volume 6122 of *Lecture Notes in Computer Science*, chapter 16, pages 140–153. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978-3-642-13600-9.
- [6] P. A. Cramer. *Controversy as an Event Category*, volume 19 of *Argumentation Library*, pages 75–137. 2011.
- [7] N. Diakopoulos, A. X. Zhang, and A. Salway. Visual analytics of media frames in online news and blogs. In *Proc. IEEE InfoVis Workshop on Text Visualization*, Oct. 2013.
- [8] R. M. Entman. Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173, 2007.
- [9] S. D. Hachohen and J. Allan. Detecting controversy on the web. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, CIKM ’13, pages 1845–1848, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2263-8.
- [10] M. Pennacchiotti and A. M. Popescu. Detecting controversies in Twitter: A first study. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, WSA ’10, pages 31–32, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [11] H. S. Rad and D. Barbosa. Identifying controversial articles in wikipedia: A comparative study. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym ’12, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1605-7.
- [12] M. Recasens, C. Danescu-Niculescu-Mizil, and D. Jurafsky. Linguistic models for analyzing and detecting biased language. In *Proc. of ACL*, pages 1650–1659, Aug. 2013.
- [13] D. A. Scheufele. Framing as a theory of media effects. *Journal of communication*, 49(1):103–122, 1999.