Media Bias, the Social Sciences, and NLP: Automating Frame Analyses to Identify Bias by Word Choice and Labeling

Felix Hamborg
Dept. of Computer and Information Science
University of Konstanz, Germany
`felix.hamborg@uni-konstanz.de`

Abstract

Media bias can strongly impact the public perception of topics reported in the news. A difficult to detect, yet powerful form of slanted news coverage is called bias by word choice and labeling (WCL). WCL bias can occur, for example, when journalists refer to the same semantic concept by using different terms that frame the concept differently and consequently may lead to different assessments by readers, such as the terms "freedom fighters" and "terrorists," or "gun rights" and "gun control." In this research project, I aim to devise methods that identify instances of WCL bias and estimate the frames they induce, e.g., not only is "terrorists" of negative polarity but also ascribes to aggression and fear. To achieve this, I plan to research methods using natural language processing and deep learning while employing models and using analysis concepts from the social sciences, where researchers have studied media bias for decades. The first results indicate the effectiveness of this interdisciplinary research approach. My vision is to devise a system that helps news readers to become aware of the differences in media coverage caused by bias.

1 Introduction

Media bias describes differences in the content or presentation of news (Hamborg et al., 2018). It is a ubiquitous phenomenon in news coverage that can have severely negative effects on individuals and society, e.g., when slanted news coverage influences voters and, in turn, also election outcomes (Alsem et al., 2008; DellaVigna and Kaplan, 2007). Potential issues of biased coverage, whether through the selection of topics or how they are covered, are compounded by the fact that in many countries only a few corporations control large parts of the media landscape–in the US, for example, six corporations control 90% of the media (Business Insider, 2014).

Even subtle changes in the words used in a news text can strongly impact readers' opinions (Papacharissi and de Fatima Oliveira, 2008; Price et al., 2005; Rugg, 1941; Schuldt et al., 2011). When referring to a semantic concept, such as a politician or other named entities (NEs), authors can label the concept, e.g., "illegal aliens," and choose from various words to refer to it, e.g., "immigrants" or "aliens." Instances of bias by word choice and labeling (WCL) frame the referred concept differently (Entman, 1993, 2007), whereby a broad spectrum of effects occurs. For example, the frame may change the polarity of the concept, i.e., positively or negatively, or the frame may emphasize specific parts of an issue, such as the economic or cultural effects of immigration (Entman, 1993).

In the social sciences, research over the past decades has resulted in comprehensive models to describe media bias as well as effective methods for the analysis of media bias, such as content analysis (McCarthy et al., 2008) and frame analysis (Entman, 1993). Because researchers need to conduct these analyses mostly manually, the analyses do not scale with the vast amount of news that is published nowadays (Hamborg et al., 2019a). In turn, such studies are always conducted for topics in the past and do not deliver insights for the current day (McCarthy et al., 2008; Oliver and Maney, 2000); this would, however, be of primary interest to people reading the news. Revealing media bias to news consumers would also help to mitigate bias effects and, for example, support them in making more informed choices (Baumer et al., 2017).

In contrast, in computational linguistics and computer science, fewer approaches systematically analyze media bias (Hamborg et al., 2019a). The models used to analyze media bias tend to be simpler (Hamborg et al., 2018; Park et al., 2009) compared to previously mentioned models. Many approaches analyze media bias from the perspective

of news consumers while neglecting both the established approaches and the comprehensive models that have already been developed in the social sciences (Evans et al., 2004; Mehler et al., 2006; Munson et al., 2013, 2009; Oelke et al., 2012; Park et al., 2009; Smith et al., 2014). Correspondingly, their results are often inconclusive or superficial, despite the approaches being technically superior.

2 Research Question, Tasks, and Contributions

To address the issues described in Section 1, I define the following research question for my Ph.D. research: *How can an automated approach identify instances of bias by word choice and labeling in a set of English news articles reporting on the same event?* To address this research question, I derive the following research tasks:

T1. Identify the strengths and weaknesses of manual and of automated methods used to identify media bias.

T2. Research NLP techniques and required datasets to address these weaknesses. To do so, use established bias models and (semi-)automate currently manual analysis methods.

T3. Implement a prototype of a media bias identification system that employs the developed methods to demonstrate the applicability of the approach in real-world news article collections. The target group of the prototype are non-expert people.

T4. Evaluate the effectiveness of the bias identification methods with a test corpus and evaluate the effectiveness of using the prototype in a user study.

Combining the expertise of the social sciences and computational linguistics appears beneficial for research on media bias. Thus, the main contribution of my Ph.D. research will be an approach that combines models and methods from multiple disciplines. On the one hand, it will leverage established models from the social sciences to describe media bias and will follow currently manual methods to analyze media bias. On the other hand, it will take advantage of scalable methods for text analysis developed and used in computational linguistics. I need to employ and extend the state-of-the-art in two closely related NLP fields (cf. Section 4): (1) cross-document coreference resolution (CDCR) as

well as (2) target-dependent sentiment classification (TSC) including "sentiment shift" and identification of framing effects and causes (see Section 4.2). I plan to embed both techniques into an approach that is inspired by the procedure of manually conducted, inductive frame analyses (cf. Section 3.1).

For the first technical contribution, a sieve-based CDCR approach was already devised that addresses characteristics of coreferences as they often occur in bias by WCL. The examples in the Abstract show that even phrases that are usually considered contrary may be coreferential in a set of articles reporting on a specific event. For the second technical contribution, i.e., to estimate how a semantic concept may be perceived by people when reading a news article, I primarily plan to devise and test neural models that I will design specifically for the task. I also plan to implement a prototype that includes visualizations to reveal the identified instances of bias by WCL to users of the system.

In the remainder of this document, I will give a brief overview of manual techniques for the analysis of bias by WCL and exemplary results from the social sciences as well as related, automated approaches (Section 3). Section 3 concludes with the current research gap, which motivates my Ph.D. research. Section 4 describes the tasks that I have already conducted as well as current and future tasks to complete my Ph.D. research. Section 5 describes a preliminary evaluation, which I already completed, as well as remaining tasks.

3 Related Work

The following summarizes an interdisciplinary literature review that I conducted as part of my Ph.D. research (T1) (Hamborg et al., 2019a).

3.1 Manual Approaches

In the social sciences, the news production process is an established model that defines nine forms of media bias and describes where these forms originate from (Baker et al., 1994; Hamborg et al., 2019a, 2018; Park et al., 2009). For example, journalists select events, sources, and from these sources the information they want to publish in a news article. While these initial selections are necessary due to the multitude of real-world events, they may also introduce bias to the resulting story. While writing an article, authors can affect readers' perception of a topic through word choice (cf. Sec-

tion 1, Baker et al., 1994; Gentzkow and Shapiro, 2006; Oelke et al., 2012). Lastly, for example, the placement and size of an article on a website determine how much attention the article will receive.

Researchers in the social sciences primarily conduct frame analyses or, more generally, content analyses to identify instances of bias by WCL (McCarthy et al., 2008; Oliver and Maney, 2000). In content analysis, researchers first define one or more analysis questions or hypotheses. Then, they gather the relevant news data, and coders systematically read the texts, annotating parts of the texts that indicate instances of bias relevant to the analysis question, e.g., phrases that change readers' perception of a specific person or topic. In inductive content analysis, coders read and annotate the texts without prior knowledge other than the analysis question. In deductive content analysis, coders must adhere to a set of coding rules defined in a codebook, which is usually created using the findings from an earlier inductive content analysis. After the coding, researchers quantify the annotated instances to lastly accept or reject their hypotheses.

Content analyses conducted for WCL bias are typically either topic-oriented or person-oriented. Annotations range from basic forms, e.g., targeted sentiment (Niven, 2002), to fine-grained "perception" categories, causes thereof, or other features, e.g., Papacharissi and de Fatima Oliveira (2008) investigated WCL in the coverage of different news outlets on topics related to terrorism. One high-level finding was that the New York Times used more dramatic tones than the Washington Post, e.g., news articles dehumanized terrorists by not ascribing any motive to terrorist attacks or use of metaphors, such as "David and Goliath." Both the Financial Times and the Guardian focused their articles on factual reporting.

3.2 (Semi-)Automated Approaches

Many automated approaches treat media bias vaguely and view it only as "differences of [news] coverage" (Park et al., 2011b), "diverse opinions" (Munson and Resnick, 2010), "different perspectives" (Hamborg et al., 2018), or "topic diversity" (Munson et al., 2009), resulting in inconclusive or superficial findings (Hamborg et al., 2019a). Only a few approaches use comprehensive bias models or focus on a specific form of media bias (cf. Section 3.1). Likewise, few approaches aim to specifically identify instances of WCL bias. For

example, Lim et al. (2018); Spinde et al. (2020b) propose to investigate words with a low document frequency in a set of news articles reporting on the same event, to find potentially biasing words that are characteristic for a single article. NewsCube 2.0 employs crowdsourcing to estimate the bias of articles reporting on a topic. The system allows users to annotate WCL in news articles collaboratively (Park et al., 2011a).

The most related, fully automated field of methods is TSC, which aims to find the connotation of a phrase regarding a given target. On news texts, however, to-date TSC methods perform poorly for at least three reasons. First, news texts have rather subtle connotations due to the expected journalistic objectivity (Gauthier, 1993; Hamborg et al., 2018). Second, to my knowledge, no news-tailored TSC approaches, dictionaries, nor annotated datasets exist; generic approaches tend to perform poorly on news texts (Balahur et al., 2010; Kaya et al., 2012; Oelke et al., 2012). Third, the one-dimensional polarity scale used by all mature TSC methods may fall short of representing complex news frames (cf. Section 1). To avoid the difficulties of highly context-dependent connotations in news texts, researchers have proposed to perform TSC only on quotes (Balahur et al., 2010) or on the readers' comments (Park et al., 2011b), which more likely contain explicit connotations. Researchers also suggested to investigate emotions induced by headlines, but they achieved mixed results (Strapparava and Mihalcea, 2007).

3.3 Research Gap

To my knowledge, there are currently no automated approaches that identify or compare instances of WCL bias, despite reliable analysis concepts used in the social sciences and automated text analysis methods in related fields, such as CDCR and TSC.

To address the difficulties due to the expected objectivity of news texts and other previously mentioned factors, I plan to follow two main ideas: first, the use of knowledge and models from sciences that have long studied media bias. Second, I expect the recent advent of word embeddings and deep learning, including neural language models, such as BERT (Devlin et al., 2018), to be strongly beneficial to the outcome of this project. The advancements in these fields have led to a performance leap in many NLP disciplines, including coreference resolution and TSC, where, e.g., in the latter

macro F1 gained from $F1_m = 63.3$ (Kiritchenko et al., 2014) to $F1_m = 75.8$ on the Twitter set (Zeng et al., 2019).

4 Methodology

Research task T2 will be the main contribution of my Ph.D. research; hence, this section focuses on completed and future tasks related to T2. Technically, addressing the research question represents two main challenges. First, resolving coreferences of semantic concepts across a set of news articles. In bias by WCL, journalists often use coreferences in a broader, sometimes even contradictory, sense than the state-of-the-art in coreference resolution and CDCR is capable of (Balahur et al., 2010; Baumer et al., 2017; Hamborg et al., 2019b). Second, classifying how actors and other semantic concepts are framed due to their mentions and their mentions' contexts, for which I will use TSC.

I plan to integrate the two tasks into the analysis shown in Figure 1 (RT3). Given a set of news articles reporting on the same event, the analysis will find subsets of articles and in-text phrases that similarly frame the concepts involved in the event. Lastly, the system will visualize the results to news consumers. Because RT3 is not directly related to NLP, it is described only briefly in Section 4.3.

4.1 Broad Cross-doc. Coreference Resolution

After the system has completed state-of-the-art preprocessing (Manning et al., 2014), the second phase in the analysis is broad CDCR, which aims to resolve coreferences as they occur in WCL bias (Hamborg et al., 2019b). The first task within this phase is *candidate extraction*. Relevant phrases containing bias by WCL commonly are noun phrases (NPs), e.g., NEs such as politicians, or verb phrases (VPs), i.e., describing an action, such as "cross the border" or "invade the country." The approach currently focuses only on NPs and extracts mentions from two sources. First, mentions from coreference chains identified by coreference resolution, and second, NPs identified by parsing.

The second task, *candidate merging*, addresses the main difficulty of broad CDCR. Journalists often use divergent terms to refer to the same semantic concept (Hamborg et al., 2019a), sometimes even terms that typically have opposing meanings, such as "intervene" vs. "invade," "coalition forces" vs. "invading forces." Such coreferences are highly context-dependent and may only be valid in a single news article or across related articles (Hamborg et al., 2019b,c). Related state-of-the-art techniques for coreference resolution capably resolve generally valid synonyms, nominal and pronominal coreferences, such as "Donald Trump," "US president," and "he." However, they cannot reliably resolve the previously mentioned, broader examples of coreferences, which often occur in bias by WCL (Hamborg et al., 2019a).

The candidate merging task uses a series of sieves, where each analyzes specific characteristics of two candidates to determine whether they should be merged (see Figure 1). For example, the first sieve merges candidates if they have similar core meanings, specifically, if the head of each candidate's representative phrase is identical (Hamborg et al., 2019b). For a given coreference chain, the representative phrase is defined as the mention that best represents the chain's meaning (Manning et al., 2014). This way, the first sieve merges cases such as "Donald Trump" and "President Trump." The second sieve merges candidates if most of their mentions are semantically similar. The sieve currently uses non-contextualized word embeddings, specifically word2vec (Mikolov et al., 2013), to vectorize each mention. Then, it calculates the unweighted mean of all vectorized mentions of a candidate. Lastly, the sieve will merge two candidates if their mean vectors are similar by cosine similarity. Analogously, the remaining sieves address specific characteristics, e.g., using word embeddings (Le and Mikolov, 2014) and clustering methods, such as affinity propagation (Frey and Dueck, 2007). More information on the approach is described by Hamborg et al. (2019b).

Future research directions for the CDCR task most importantly include extending the capabilities of the approach and improving its performance. For the former, we want to investigate how coreferential mentions of activities (VPs) can be resolved. To improve the CDCR performance, we plan to devise a method that uses a language model to resolve coreferential mentions. For example, BERT increased the performance on single-document coreference resolution from F1=73.0 to F1=77.1. Using SpanBERT, a pre-training method focused on spans rather than tokens, the performance is increased to F1=79.6 (Joshi et al., 2019). We expect that using a language model can yield similar improvements for CDCR.
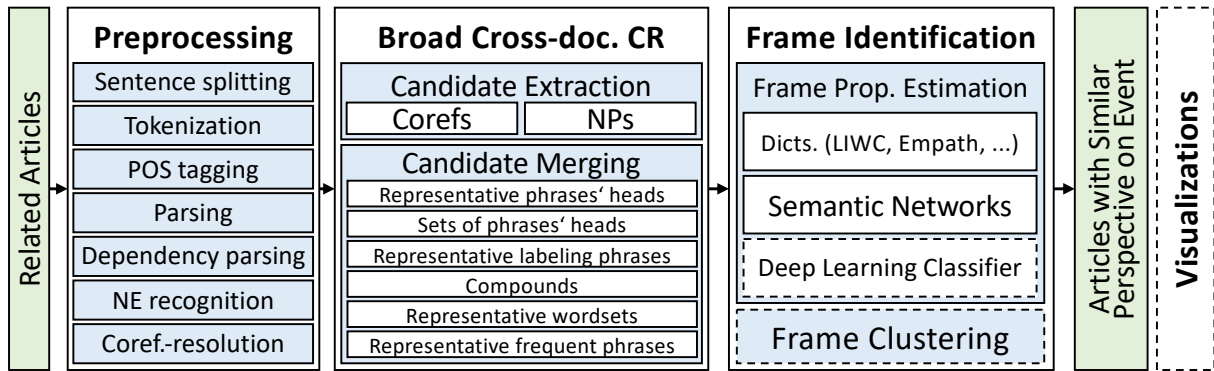
Figure 1: Shown is the plan for the three-phase analysis pipeline as it preprocesses news articles reporting on the same event, resolves coreferential mentions of semantic concepts across documents, and groups articles framing these concepts similarly. Source: (Hamborg et al., 2019b)

4.2 Frame Identification

Approaches aiming to estimate how semantic concepts are perceived, e.g., in the closely related field of TSC by classifying the concepts' polarity, or, more broadly, approaches to identify bias, traditionally employ manually created dictionaries or manually engineered features for machine learning (ML). Such approaches can achieve high performances in various domains, e.g., Recasens et al. (2013) propose an approach that capably identifies single bias-words in Wikipedia articles by using dictionaries and further, non-complex features.

In news texts, however, such approaches fall short. Since neutral language is expected (cf. Section 3), token-based and ML methods fail to catch the "meaning between the lines" (Hamborg et al., 2019a,b; Balahur et al., 2010; Godbole et al., 2007). Yet, recent NLP advancements, most importantly language models, have proven to be very effective in the news domain as in various other domains and tasks (see Section 3.3).

I plan to devise a neural model that will, in part, be inspired by state-of-the-art TSC approaches such as LCF-BERT (Zeng et al., 2019) and domain-adapted SPC-BERT (Rietzler et al., 2019), with three main differences. First, the model will need to consider characteristics specific to news articles. For example, in news articles, sentiment may more strongly depend on global context compared to TSC prime domains, e.g., because the latter are typically shorter texts (Adhikari et al., 2019).

Second, besides "absolute" sentiment polarity, the model needs to consider the "sentiment shift" induced by the context of a target mention. For example, while TSC traditionally focuses on the event's or text's sentiment regarding a target (cf.

"text-level" as defined by Balahur et al. (2010)), bias by WCL is concerned explicitly with the language, e.g., words, used in the sentence. So, given a target mention, I am interested in whether the mention or its context sway the perception more positively or negatively, also in relation to the sentiment at event- or text-level (Balahur et al., 2010).

Third, for an identified non-neutral polarity, the approach should be able to find in-text causes and potential effects thereof. Causes include the use of emotional words, loaded language, or aggressive repetition of specific facts. Effects include particularly how the target is framed (cf. "frame properties" as defined by Hamborg et al. (2019b) or "frames types" by Card et al. (2015)). Resolving the dependencies of a target and its context is an issue that is subject of current TSC research (Zeng et al., 2019; Rietzler et al., 2019), which I expect to be important in the proposed project as well.

4.3 System and Visualization

A system will integrate the previously described analysis workflow and will visualize the results to non-expert users (RT3). I devised visualizations that are similar to UIs of popular news aggregators, such as Google News, and bias-aware aggregators, such as AllSides. In contrast to these, the system will be able to identify in-text instances of bias (Hamborg et al., 2017, 2020; Spinde et al., 2020a). Hence, the system will not only give a bias-aware overview of current topics but also will have a visualization for single articles, which will highlight identified instances of WCL bias.

For research and evaluation of the previously described system and its analysis methods, I currently use the datasets AllSides (Chen et al., 2018), NewsWCL50 (Hamborg et al., 2019c), and PO-

LUSA (Gebhard and Hamborg, 2020), which have high diversity concerning outlets' political slant.

I plan to publish the code of the system and methods. Due to the system's modularity, researchers can extend it to support further forms of bias, e.g., commission and omission of information or picture selection (Torres, 2018; Hamborg et al., 2019a).

5 Evaluation

I conducted preliminary evaluations of the two main methods described in Section 4 (RT4). To measure the CDCR performance on broad coreferences as they occur in WCL bias (Section 4.1), I created a test dataset named NewsWCL50. The dataset was created by manually annotating coreferential mentions of persons, actions, and also vaguely defined, abstract concepts across 50 news articles (Hamborg et al., 2019b). The evaluation seems to confirm the research direction for this task. The approach currently achieves $F1 = 45.7$, or 84.4 if evaluated only on technically feasible annotations, compared to 29.8, or 42.1, respectively, achieved by the best baseline. Technically feasible refers to only comparing to annotations that the approach theoretically should be able to resolve, e.g., currently only NPs while excluding VPs.

A future evaluation will include a comparison to state-of-the-art CDCR methods (Barhom et al., 2019; Intel AI Lab, 2018). For improved soundness, we plan to create a second dataset similar to the NewsWCL50 dataset but with more coders and more articles. To do so, we will crowdsource the annotations of concept mentions on MTurk and use an improved codebook. The improvements will address issues of NewsWCL50's codebook, e.g., by making annotation types less ambiguous (Hamborg et al., 2019b). Further, we plan to use two additional datasets: ECB+ (Cybulska and Vossen, 2014) and NIdent (Recasens et al., 2012). Both datasets are commonly used to evaluate CDCR approaches and contain cross-document coreferences.

To evaluate the second task, frame identification, I plan to create a comprehensive training and test set for the TSC method described in Section 4.2. I already created a preliminary dataset of 3000 sentences, each including a target mention and a sentiment label agreed on by three coders. The dataset was created analogously to established TSC datasets (Dong et al., 2014; Pontiki et al., 2014; Nakov et al., 2016; Rosenthal et al., 2017).

Preliminary results seem to indicate that TSC on the news domain is in part more difficult than on TSC prime domains, such as product reviews, where authors often express their opinion explicitly. State-of-the-art TSC achieves average recall $AvgRec = 70.0$ on news articles, whereas performances on common TSC test datasets range from $AvgRec = 75.6$ (Twitter dataset) to 82.2 (Restaurant). Other baselines, e.g., using dictionaries and semantic networks, such as ConceptNet, perform very poorly ($F1 < 15.0$), which seems to confirm that token-based approaches fail to catch the subtlety common to WCL bias.

Finally, we plan to evaluate the system's effectiveness regarding visualization of the identified biases to non-expert users. An already conducted pre-study confirmed the study design (Spinde et al., 2020a). I will revisit this task once the classification methods described in Section 4 can be used within the study.

6 Conclusion and Implications

In summary, both everyday news consumers, as well as researchers in the social sciences, could benefit strongly from the automated identification of bias by word choice and labeling (WCL) in news articles. Devising suitable methods to resolve broad coreferences across news articles reporting on the same event and estimating the frames of the found instances of WCL bias are at the heart of this research project. One primary result of the project will be the first automated approach capable of identifying instances of bias by WCL in a set of news articles reporting on the same event or topic.

My vision is that at a later point in time, such methods might be integrated into popular news aggregators, such as Google News, helping news readers to explore and understand media bias through their daily news consumption. Also, I think that these methods could be integrated into the analysis workflow of content analyses and frame analyses, helping to automate further these currently mostly manual and thus time-consuming analysis concepts prevalent in the social sciences.

References

Ashutosh Adhikari, Achyudh Ram, Raphael Tang, and Jimmy Lin. 2019. DocBERT: BERT for Document Classification.

Karel Jan Alsem, Steven Brakman, Lex Hoogduin, and Gerard Kuper. 2008. The impact of newspapers on consumer confidence: Does spin bias exist? *Applied Economics*, 40(5):531–539.

Brent H Baker, Tim Graham, and Steve Kaminsky. 1994. *How to identify, expose & correct liberal media bias*. Media Research Center, Alexandria, VA.

Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2010. Sentiment analysis in the news. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation ({LREC}'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Shany Barhom, Vered Shwartz, Alon Eirew, Michael Bugert, Nils Reimers, and Ido Dagan. 2019. Revisiting Joint Modeling of Cross-document Entity and Event Coreference Resolution. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4179–4189, Stroudsburg, PA, USA. Association for Computational Linguistics.

Eric P.S. Baumer, Francesca Polletta, Nicole Pierski, and Geri K. Gay. 2017. A Simple Intervention to Reduce Framing Effects in Perceptions of Global Climate Change. *Environmental Communication*.

Business Insider. 2014. These 6 Corporations Control 90% Of The Media In America.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The Media Frames Corpus: Annotations of Frames Across Issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Stroudsburg, PA, USA. Association for Computational Linguistics.

Wei-Fan Chen, Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2018. Learning to Flip the Bias of News Headlines. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 79–88, Stroudsburg, PA, USA. Association for Computational Linguistics.

Agata Cybulska and Piek Vossen. 2014. Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4545–4552, Reykjavik, Iceland. European Language Resources Association (ELRA).

Stefano DellaVigna and Ethan Kaplan. 2007. The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv: 1810.04805*.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive Recursive Neural Network for target-dependent Twitter sentiment classification. In *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014*, pages 49–54, Baltimore, MD, USA.

Robert M Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Robert M Entman. 2007. Framing Bias: Media in the Distribution of Power. *Journal of Communication*, 57(1):163–173.

David Kirk Evans, Judith L. Klavans, and Kathleen R. McKeown. 2004. Columbia Newsblaster. In *Demonstration Papers at HLT-NAACL 2004 on XX - HLT-NAACL '04*, pages 1–4, Morristown, NJ, USA. Association for Computational Linguistics.

Brendan J. Frey and Delbert Dueck. 2007. Clustering by Passing Messages Between Data Points. *Science*, 315(5814):972–976.

Gilles Gauthier. 1993. In Defence of a Supposedly Outdated Notion: The Range of Application of Journalistic Objectivity. *Canadian Journal of Communication*, 18(4):497.

Lukas Gebhard and Felix Hamborg. 2020. The PO-LUSA Dataset: 0.9M Political News Articles Balanced by Time and Outlet Popularity. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 1–2, Wuhan, CN.

Matthew Gentzkow and Jesse M. Shapiro. 2006. Media Bias and Reputation. *Journal of Political Economy*, 114(2):280–316.

Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. 2007. Large-Scale Sentiment Analysis for News and Blogs. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, volume 7, pages 219–222, Boulder, CO, USA.

Felix Hamborg, Karsten Donnay, and Bela Gipp. 2019a. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4):391–415.

Felix Hamborg, Norman Meuschke, Akiko Aizawa, and Bela Gipp. 2017. Identification and Analysis of Media Bias in News Articles. In *Proceedings of the 15th International Symposium of Information*

Science, pages 224–236, Berlin, DE. Verlag Werner Hülsbusch.

Felix Hamborg, Norman Meuschke, and Bela Gipp. 2018. Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries*.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2019b. Automated Identification of Media Bias by Word Choice and Labeling in News Articles. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 196–205, Urbana-Champaign, IL, USA. IEEE.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2019c. Illegal Aliens or Undocumented Immigrants? Towards the Automated Identification of Bias by Word Choice and Labeling. In *Proceedings of the iConference 2019*, pages 179–187. Springer, Cham, Washington, DC, USA.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2020. Newsalyze: Enabling News Consumers to Understand Media Bias. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 1–2, Wuhan, CN.

Intel AI Lab. 2018. NLP Architect.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2019. SpanBERT: Improving Pre-training by Representing and Predicting Spans.

Mesut Kaya, Guven Fidan, and Ismail H Toroslu. 2012. Sentiment Analysis of Turkish Political News. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pages 174–180. IEEE Computer Society, IEEE.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland. Association for Computational Linguistics.

Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. *International Conference on Machine Learning - ICML 2014*, 32.

Sora Lim, Adam Jatowt, and Masatoshi Yoshikawa. 2018. Towards Bias Inducing Word Detection by Linguistic Cue Analysis in News Articles.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Stroudsburg, PA, USA. Association for Computational Linguistics.

John McCarthy, Larissa Titarenko, Clark McPhail, Patrick Rafail, and Boguslaw Augustyn. 2008. Assessing stability in the patterns of selection bias in newspaper coverage of protest during the transition from communism in Belarus. *Mobilization: An International Quarterly*, 13(2):127–146.

Andrew Mehler, Yunfan Bao, X. Li, Y. Wang, and Steven Skiena. 2006. Spatial Analysis of News Sources. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):765–772.

Tomas Mikolov, Greg Corrado, Kai Chen, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*.

Sean A. Munson, Stephanie Y. Lee, and Paul Resnick. 2013. Encouraging reading of diverse political viewpoints with a browser widget. In *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*.

Sean A Munson and Paul Resnick. 2010. Presenting diverse political opinions. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, pages 1457–1466, New York, New York, USA. ACM, ACM Press.

Sean A Munson, Daniel Xiaodan Zhou, and Paul Resnick. 2009. Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators. In *ICWSM*.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. SemEval-2016 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1–18, San Diego, CA, USA. Association for Computational Linguistics.

David Niven. 2002. *Tilt?: The search for media bias*. Praeger.

Daniela Oelke, Benno Geißelmann, and Daniel A Keim. 2012. Visual Analysis of Explicit Opinion and News Bias in German Soccer Articles. In *EuroVis Workshop on Visual Analytics*, Vienna, Austria.

Pamela E. Oliver and Gregory M. Maney. 2000. Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions. *American Journal of Sociology*, 106(2):463–505.

Zizi Papacharissi and Maria de Fatima Oliveira. 2008. News Frames Terrorism: A Comparative Analysis of Frames Employed in Terrorism Coverage in U.S. and U.K. Newspapers. *The International Journal of Press/Politics*, 13(1):52–74.

Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. NewsCube. In *Proceedings of the 27th international conference on Human factors*

in computing systems - CHI 09, page 443, New York, New York, USA. ACM Press.

Souneil Park, Minsam Ko, Jungwoo Kim, H Choi, and Junehwa Song. 2011a. NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation. In *Workshop on Social Recommender Systems*.

Souneil Park, Minsam Ko, Jungwoo Kim, Ying Liu, and Junehwa Song. 2011b. The politics of comments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11*, pages 113–122, New York, New York, USA. ACM, ACM Press.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland.

Vincent Price, Lilach Nir, and Joseph N. Cappella. 2005. Framing public discussion of gay civil unions.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic Models for Analyzing and Detecting Biased Language. In *Proceedings of the 51st Annual Meeting on Association for Computational Linguistics*, pages 1650–1659, Sofia, BG. Association for Computational Linguistics.

Marta Recasens, M. Antonia Marti, and Constantin Orasan. 2012. Annotating Near-Identity from Coreference Disagreements. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 165–172, Istanbul, Turkey. European Language Resources Association (ELRA).

Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2019. Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification. *arXiv preprint arXiv:1908.11860*.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 Task 4: Sentiment Analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.

Donald Rugg. 1941. Experiments in Wording Questions: II. *Public Opinion Quarterly*.

Jonathon P. Schuldt, Sara H. Konrath, and Norbert Schwarz. 2011. Global warming or climate change? Whether the planet is warming depends on question wording. *Public Opinion Quarterly*.

Alison Smith, Timothy Hawes, and Meredith Myers. 2014. Hiérarchie: Interactive Visualization for Hierarchical Topic Models. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 71–78.

Timo Spinde, Felix Hamborg, Angelica Becerra, Karsten Donnay, and Bela Gipp. 2020a. Enabling News Consumers to View and Understand Biased News Coverage: A Study on the Perception and Visualization of Media Bias. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 1–4, Wuhan, CN.

Timo Spinde, Felix Hamborg, and Bela Gipp. 2020b. An Integrated Approach to Detect Media Bias in German News Articles. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 1–2, Wuhan, CN.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74, Prague, Czech Republic. Association for Computational Linguistics.

Michelle Torres. 2018. Give me the full picture: Using computer vision to understand visual frames and political communication.

Biqing Zeng, Heng Yang, Ruyang Xu, Wu Zhou, and Xuli Han. 2019. LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification. *Applied Sciences*, 9(16):1–22.