Illegal Aliens or Undocumented Immigrants? Towards the Automated Identification of Bias by Word Choice and Labeling

Felix Hamborg, Anastasia Zhukova, Bela Gipp

University of Konstanz, Germany
{firstname.lastname}@uni-konstanz.de

Abstract. Media bias, i.e., slanted news coverage, can strongly impact the public perception of topics reported in the news. While the analysis of media bias has recently gained attention in computer science, the automated methods and results tend to be simple when compared to approaches and results in the social sciences, where researchers have studied media bias for decades. We propose Newsalyze, a work-in-progress prototype that imitates a manual analysis concept for media bias established in the social sciences. Newsalyze aims to find instances of bias by word choice and labeling in a set of news articles reporting on the same event. Bias by word choice and labeling (WCL) occurs when journalists use different phrases to refer to the same semantic concept, e.g., actors or actions. This way, instances of bias by WCL can induce strongly divergent emotional responses from readers, such as the terms "illegal aliens" vs. "undocumented immigrants." We describe two critical tasks of the analysis workflow, finding and mapping such phrases, and estimating their effects on readers. For both tasks, we also present first results, which indicate the effectiveness of exploiting methods and models from the social sciences in an automated approach.

Keywords: Media bias, news slant, news bias, content analysis, frame analysis.

1 Introduction

Media bias describes differences in the content or presentation of news [23]. It is an ubiquitous phenomenon in news coverage that can have severely negative effects on individuals and society [23], for example when slanted news coverage influences voters and, in turn, also election outcomes [1, 11]. Potential issues of one-sided coverage, whether through 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, only six corporations control 90% of the media [6].

Subtle changes in the words used in a news text can significantly impact opinions [43, 47, 49, 50]. When referring to a semantic concept, such as a politician or generally named entities, 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 [13, 14, 39], whereby a broad spectrum of effects occurs [22], e.g., 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 economical or cultural effects of immigration [13].

In the social sciences, research over the past decades has developed comprehensive models to describe media bias as well as effective methods for the analysis of media bias, such as the *content analysis* [32] and the *frame analysis* [13]. Because researchers need to conduct these analyses mostly manually, the analyses do not scale with the vast amount of news that is published nowadays. In turn, such studies are always conducted for topics in the past, and do not deliver insights for the current day (cf. [32, 42]), which would, however, be of primary interest to regular news consumers. Revealing media bias to news consumer would also help to mitigate bias effects, and, for example, support news consumer in making more informed choices [22].

In contrast, in computer science, few approaches systematically analyze media bias. The models used to analyze media bias in computer science tend to be simplistic (cf. [23, 26, 36, 37, 44, 51]) compared to models established in the social sciences; most approaches analyze media bias from the perspective of every-day news readers while neglecting both the established approaches and the comprehensive models that have already been developed in the social sciences (cf. [15, 33, 36, 37, 41, 44, 51]). Correspondingly, their results are often inconclusive or superficial, despite the approaches being technically promising. To address these issues, we define the research question:

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, and enable every-day news consumers to explore these instances?

We propose a cross-disciplinary approach that exploits the established models from the social sciences to describe and methods to analyze media bias, while taking advantage of the fast, scalable methods for text analysis developed and used in computer science (Section 3). Our approach imitates the process of an inductive frame analysis, and uses state-of-the-art natural language processing (NLP) methods to identify and map bias inducing coreferences (currently only noun phrases (NPs), i.e., phrases referring to the same semantic concept. To estimate the effects of such coreferences on readers, we use psychometric dictionaries devised in psychology and linguistics. Further contributions are a brief overview of techniques for the analysis of bias by WCL and exemplary results from the social sciences and related approaches from computer science (Section 2), and first results demonstrating the effectiveness of our cross-disciplinary approach (Section 3). We conclude our paper with future work for the prototype (Section 4).

2 Related Work

In the social sciences, the *news production and consumption process* is an established model that defines nine forms of media bias, and describes where these forms originate from [3, 23, 44]. For example, first journalists *select events, sources*, and from these sources the *information* they want to publish in a news article. These initial selection processes introduce a bias into the resulting news story. While writing an article, journalists can affect the reader's perception of a topic through *word choice and labeling* as described in Section 1 [3, 19, 41]. Lastly, the *placement* and *size* of an article within a newspaper or 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, and investigate their effects on individuals or societies [32, 42].[1] In a content analysis, researchers first define analysis questions or hypotheses. Then, they gather the relevant news texts, and coders read the texts, annotating parts of the texts that indicate instances of media bias relevant to the analysis questions, e.g., phrases that change the readers' perception of a specific person or topic. In an *inductive* content analysis, coders read and annotate the texts without prior knowledge other than the analysis question. In a *deductive* content analysis, coders adhere to a set of coding rules defined in a code book, which researchers usually create using the findings from an inductive content analysis conducted prior to the deductive analysis. After the coding, researchers use the annotated findings, for example, to accept or reject their hypotheses.

The content analyses conducted for bias by WCL are typically either *topic-oriented* or *person-oriented.* For example, Papacharissi and Oliveira investigated WCL in the coverage of different news outlets on topics related to terrorism [43]. 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 usage of metaphors, such as "David and Goliath" [43]. Both the Financial Times and the Guardian focused their news articles on factual reporting. Another study analyzed whether articles portrayed Bill Clinton, the U.S. president at that time, positively, neutrally, or negatively [40].

Most automated approaches treat media bias vaguely, and view it only as "differences of [news] coverage" [46], "diverse opinions" [38], or "topic diversity" [37], resulting in inconclusive or superficial findings [21]. Few approaches use the bias models from the social sciences and focus on a specific form of media bias. Likewise, few approaches specifically aim to identify instances of bias by WCL. Lim et al. 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 [30]. 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 [45]. A closely related, fully automated field of methods is *sentiment analysis*, which aims to find the connotation of a phrase. On news texts, however, sentiment analysis performs poorly for three reasons. First, news texts have rather subtle connotations due to the journalistic objectivity [18, 23]. Second, no sentiment dictionary exists that is specifically designed for news texts, and generic dictionaries tend to perform poorly on news texts (cf. [4, 28, 41]). Third, the one-dimensional positive-negative scale used by all mature sentiment analyzers likely falls short of representing the complexity of news articles [41]. To avoid the difficulties of highly context-dependent sentiment connotations in news articles, researchers have proposed approaches to perform sentiment analysis specifically on quotes [4] or on the comments of readers [46], which more likely contain an explicit statement of sentiment. First research projects suggested

[1] The paragraphs about manual and automated approaches have been adapted partially from [21].

to investigate *emotions* induced by headlines but they achieved mixed results [54]. Other approaches use dictionaries to find bias words in Wikipedia articles [48] and news articles [5]. Both approaches achieve an accuracy close to human coders, but do not estimate the effects of the found words on readers.

In conclusion, there is currently no automated approach that enables users to view instances of bias by WCL in news coverage of the current day, despite the reliable analysis concepts developed and used in the social sciences, and fast, scalable text analysis methods developed in computer science and computational linguistics.

3 Identification of Bias by Word Choice and Labeling

Newsalyze is a research prototype that aims to find groups of articles that *frame* an event similarly, i.e., report similarly on the named entities (NEs) and other semantic concepts involved in the event. Therefore, Newsalyze implements a three-tasks analysis pipeline as depicted in Fig. 1. From a set of articles reporting on the same event, Newsalyze first performs state-of-the-art NLP *preprocessing*. The second task, *frame device analysis*, finds so called *frame devices* [8], i.e., in our project phrases referring to any concept (candidate extraction), and then aligns all phrases referring to the same concept from all articles (candidate alignment). Our prototype currently analyzes noun phrases (NPs). The third task, *frame identification*, estimates the effect of such phrases on readers (Effect on Readers (EoR) estimation), and finally clusters articles that have a similar EoR of aligned phrases (frame clustering). The output of the system are groups of articles framing the event similarly, which the system *visualizes* to users finally.

The two main challenges in automatically identifying instances of bias by WCL are the candidate alignment and the EoR estimation, which we describe in more detail.

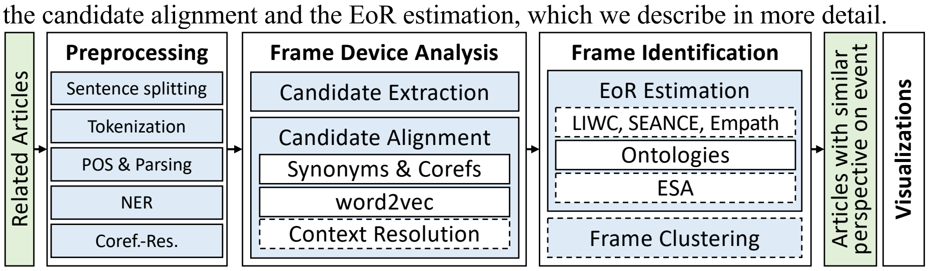

Fig. 1. The three-tasks analysis pipeline preprocesses news articles, extracts and aligns phrases referring to the same semantic concepts, and groups articles reporting similarly on these concepts.

3.1 Candidate Alignment

The first task is to align coreferences across multiple articles, commonly called *cross-document coreference resolution*, a task that current NLP methods cannot reliably perform for coreferences as they occur in bias by WCL in news articles. Current NLP methods, such as named entity linking (NEL), NE recognition (NER) and disambiguation (NERD), and coreference resolution capably identify synonyms of the same NE, such as 'Mr. Trump' and 'US President' (precision $p \approx 0.8$ [7]), and pronominal and nominal coreferences, such as 'he' and 'Donald J. Trump' ($p \approx 0.8$ [9]).

The automated alignment of WCL candidates, however, is more challenging because often journalists refer to the same concept in a broader sense than currently addressed by the previously mentioned coreference resolution methods. Instead, such coreferences are highly dependent on the context, may only be valid in a single article or across related articles, or are only meaningful in conjunction with an attribute, e.g., in articles reporting on the end of DACA in 2018 the terms 'illegal aliens' and 'undocumented immigrants' referred to the people that were protected by DACA [29].

To align WCL candidates, we currently use word embeddings produced by word2vec with the generic Google News model (300M words) [34]. Specifically, in the candidate extraction we extract all NPs, such as "undocumented immigrants," and mentions of coreference chains. Then, we use affinity propagation [17] on the Euclidean distance in the word2vec space to align coreferential NPs. First results depicted in Table 1 indicate the suitability of the approach to align such coreferences, e.g., our approach was able to align the bias by WCL instances "undocumented worker" and "illegal immigrant" across multiple articles of the DACA topic. For each topic, we collected articles published in the year shown in Table 1 from major news outlets representing the whole political spectrum from the US (DACA and Denuclearization) and UK (Brexit). Such coreferences cannot be resolved by neither coreference resolution nor NER.

Table 1. Automatically aligned coreferences for exemplary topics.

Topic	# articles	Aligned coreferences
DACA US, 2017	25	immigrant(s), migrant(s), illegal immigrant(s), undocumented immigrants, Latino immigrants, illegals, undocumented workers, sympathetic group of immigrants, …
Denucleari- zation PRK, 2018	25	American military presence, unilateral US military operation, US mere presence, US military action, US military installations, US military presence, US military threat, …
Brexit EU, 2016	35	Brexit negotiations, EU divorce negotiations, exit negotiations, discussions, negotiations between UK and EU, negotiations over terms of divorce, …

3.2 Estimation of the Effects on Readers

To estimate the EoR of coreferences, only considering their sentiment would not be sufficient, mainly due to the complexity of news topics, and also due to the subtlety of connotations motivated by the journalistic objectivity (see Section 2). Thus, we analyze *frame properties*, which we define as properties that make up a frame induced by a phrase, including emotions [31], polarity (cf. [2]), and topic-specific properties common in frame analyses (see Section 2), such as in person-oriented news competence, honesty, wisdom, and empathy (cf. [43]); and on a broader scale also topical categories, such as finance, economic, and culture (cf. [16]).

The current prototype estimates the EoR by comparing terms in documents to a set of seed words representing frame properties. We derived the following frame properties from an inductive content analysis, which we conducted on the topics shown in Table

1: aggression, honesty, competence, authority, confidence, sympathy, and their antonyms. We also add seed words representing six basic emotions [12] to the set: anger, disgust, fear, happiness, sadness, and surprise. Our initial findings on the topics from Table 1 indicate the effectiveness of our approach, e.g., in the DACA topic the most frequent frame properties ascribed by the right news website Fox News to US President Trump are mostly positive, such as honesty, sympathy, happiness, whereas the left New York Times used mainly negative properties, such as anger, fear, and sadness. Our findings are conformal with manually conducted studies on the ideological placement of the news outlets (cf. [20, 35]).

4 Conclusion and Future Work

Newsalyze is a work-in-progress prototype that aims to automatically identify instances of bias by word choice and labeling in a set of news articles reporting on the same event. In this paper, we describe the key concepts of the two fundamental tasks of the analysis workflow, i.e., the (1) alignment of context-dependent coreferences and the (2) estimation of the effects of the aligned coreferences on readers (EoR). In the first task, we currently use parsing and word2vec to find phrases referring to the same semantic concept. This way, Newsalyze finds and aligns context-specific coreferences, such as "undocumented worker" and "illegal immigrant" in the context of an immigration topic, that cannot be found by generic methods such as coreference resolution or synonym resolution. In the second task, we use a predefined set of frame properties, such as aggression and competence, represented in ConceptNet to analyze the EoR of phrases.

While the first results generally indicate the usefulness of the approach, we propose the following improvements. For candidate alignment, we want to improve the results by training a custom word2vec model on a current news dataset, for instance from the commoncrawl archive using a news crawler [27]. Besides the word2vec-based alignment approach, we plan to devise a second, syntax-based approach, which analyzes the relations between extracted candidates, such as the constituents of a sentence, particularly subject-predicate-object triples, e.g., using OpenIE [2], and event descriptors, such as the journalistic 5W phrases [24, 25], which describe the main event of news articles. The conceptual idea is that if, for example, supposedly different subjects perform the same action in related news articles, the subjects will likely refer to the same actor. We also want to investigate the extraction and alignment of non-NP coreferences, e.g., when activities are described differently, e.g., "invade" or "cross border."

To estimate the EoR, we also want to investigate the use of dictionaries, such as LIWC [55], SEANCE [10], Empath [16], categories from the General Inquirer [52], and dictionaries of bias-inducing phrases (cf. [5, 48]). We think, however, that our current approach is better at estimating the effects of new terms, since ConceptNet is updated regularly from Wikipedia and other sources [53]. Finally, we need to cluster articles similarly framing an event, e.g., articles with similarly framed coreferences.

Lastly, we need to visualize the results of the automated analysis to every-day news consumers. A news topic view used in the bias-aware news analysis could show phrases containing the most contrastive cases of bias by word choice and labeling [23].

Acknowledgements

This work was partially supported by the Carl Zeiss Foundation and the Zukunftskolleg program of the University of Konstanz. We thank the anonymous reviewers for their valuable comments.

References

[1] Alsem, K.J. et al. 2008. The impact of newspapers on consumer confidence: Does spin bias exist? *Applied Economics*. 40, 5 (2008), 531–539. DOI:https://doi.org/10.1080/00036840600707100.

[2] Angeli, G. et al. 2015. Leveraging Linguistic Structure For Open Domain Information Extraction. *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)* (2015), 344–354.

[3] Baker, B.H. et al. 1994. *How to identify, expose & correct liberal media bias*. Media Research Center Alexandria, VA.

[4] Balahur, A. et al. 2013. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*. (2013).

[5] Baumer, E.P.S. et al. 2015. Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News. *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL* (Denver, Colorado, USA, 2015), 1472–1482.

[6] Business Insider 2014. These 6 Corporations Control 90% Of The Media In America. *Lutz, Ashley*.

[7] Chang, A.X. et al. A comparison of Named-Entity Disambiguation and Word Sense Disambiguation. *LREC* 860–867.

[8] Chong, D. and Druckman, J.N. 2007. Framing Theory. *Annual Review of Political Science*. 10, 1 (2007), 103–126. DOI:https://doi.org/10.1146/annurev.polisci.10.072805.103054.

[9] Clark, K. and Manning, C.D. 2016. Deep Reinforcement Learning for Mention-Ranking Coreference Models. *Empirical Methods on Natural Language Processing* (2016).

[10] Crossley, S.A. et al. 2017. Sentiment Analysis and Social Cognition Engine (SEANCE): An automatic tool for sentiment, social cognition, and social-order analysis. *Behavior Research Methods*. 49, 3 (2017), 803–821. DOI:https://doi.org/10.3758/s13428-016-0743-z.

[11] DellaVigna, S. and Kaplan, E. 2007. The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*. 122, 3 (2007), 1187–1234. DOI:https://doi.org/10.3386/w12169.

[12] Ekman, P. 1992. An Argument for Basic Emotions. *Cognition and Emotion*. (1992). DOI:https://doi.org/10.1080/02699939208411068.

[13] Entman, R.M. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*. 43, 4 (1993), 51–58.

[14] Entman, R.M. 2007. Framing bias: Media in the distribution of power. *Journal of communication*. 57, 1 (2007), 163–173.

[15] Evans, D.K. et al. 2004. Columbia newsblaster: multilingual news summarization on the Web. *Demonstration Papers at HLT-NAACL 2004* (2004), 1–4.

[16] Fast, E. et al. 2016. Empath: Understanding Topic Signals in Large-Scale Text. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI'16)* (San Jose, California, USA, 2016), 4647–4657.

[17] Frey, B.J. and Dueck, D. 2007. Clustering by passing messages between data points. *Science*. 315, 5814 (2007), 972–976. DOI:https://doi.org/10.1126/science.1136800.

[18] Gauthier, G. 1993. In defence of a supposedly outdated notion: The range of application of journalistic objectivity. *Canadian Journal of communication*. 18, 4 (1993), 497.

[19] Gentzkow, M. and Shapiro, J. 2006. Media bias and reputation. *Journal of Political Economy*. 114, 2 (2006), 280–316. DOI:https://doi.org/10.1086/499414.

[20] Groseclose, T. and Milyo, J. 2005. A measure of media bias. *The Quarterly Journal of Economics*. 120, 4 (2005), 1191–1237. DOI:https://doi.org/10.1162/003355305775097542.

[21] Hamborg, F. et al. 2018. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*. (2018). DOI:https://doi.org/10.1007/s00799-018-0261-y.

[22] Hamborg, F. et al. 2018. Automated Identification of Media Bias in News Articles: An Interdisciplinary Literature Review. *International Journal on Digital Libraries*. (2018). DOI:https://doi.org/10.1007/s00799-018-0261-y.

[23] Hamborg, F. et al. 2018. Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries*. (2018). DOI:https://doi.org/10.1007/s00799-018-0239-9.

[24] Hamborg, F. et al. 2018. Extraction of Main Event Descriptors from News Articles by Answering the Journalistic Five W and One H Questions. *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL)* (Fort Worth, Texas, USA, 2018), 339–340.

[25] Hamborg, F. et al. 2018. Giveme5W: Main Event Retrieval from News Articles by Extraction of the Five Journalistic W Questions. *Proceedings of the iConference 2018* (Sheffield, UK, 2018), 355–356.

[26] Hamborg, F. et al. 2017. Identification and Analysis of Media Bias in News Articles. *Proceedings of the 15th International Symposium of Information Science* (2017), 224–236.

[27] Hamborg, F. et al. 2017. news-please: A Generic News Crawler and Extractor. *Proceedings of the 15th International Symposium of Information Science* (2017), 218–223.

[28] Kaya, M. et al. 2012. Sentiment analysis of turkish political news. *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01* (2012), 174–180.

[29] Lee, Y. 2006. To Dream or Not to Dream: A Cost-Benefit Analysis of the Development, Relief, and Education for Alien Minors (Dream) Act. *Cornell Journal of Law and Public Policy*. 16, 1 (2006), 231.

[30] Lim, S. et al. 2018. Towards Bias Inducing Word Detection by Linguistic Cue Analysis in News Articles. (2018).

[31] Lin, K.H.-Y. et al. 2007. What emotions do news articles trigger in their readers?

Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07 (2007), 733.

[32] McCarthy, J. et al. 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 (2008), 127–146.

[33] Mehler, A. et al. 2006. Spatial analysis of news sources. *IEEE Transactions on Visualization and Computer Graphics* (2006).

[34] Mikolov, T. et al. 2013. Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations (ICLR 2013)*. (2013). DOI:https://doi.org/10.1162/153244303322533223.

[35] Mitchell, A. et al. 2014. Political Polarization and Media Habits. *Pew Research Center*. (2014).

[36] Munson, S.A. et al. 2013. Encouraging Reading of Diverse Political Viewpoints with a Browser Widget. *ICWSM* (2013).

[37] Munson, S.A. et al. 2009. Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators. *ICWSM* (2009).

[38] Munson, S.A. and Resnick, P. 2010. Presenting diverse political opinions: how and how much. *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), 1457–1466.

[39] News Bias Explored - The art of reading the news: 2014. *http://umich.edu/~newsbias/*. Accessed: 2018-08-01.

[40] Niven, D. 2002. *Tilt?: The search for media bias*. Greenwood Publishing Group.

[41] Oelke, D. et al. 2012. Visual Analysis of Explicit Opinion and News Bias in German Soccer Articles. *EuroVis Workshop on Visual Analytics* (Vienna, Austria, 2012).

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

[43] Papacharissi, Z. and de Fatima Oliveira, M. 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 (2008), 52–74. DOI:https://doi.org/10.1177/1940161207312676.

[44] Park, S. et al. 2009. NewsCube: Delivering Multiple Aspects of News to Mitigate Media Bias. *Proceedings of CHI '09, the SIGCHI Conference on Human Factors in Computing Systems*. (2009), 443–453. DOI:https://doi.org/10.1145/1518701.1518772.

[45] Park, S. et al. 2011. NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation. *Workshop on Social Recommender Systems* (2011).

[46] Park, S. et al. 2011. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. *Proceedings of the ACM 2011 conference on Computer supported cooperative work* (2011), 113–122.

[47] Price, V. et al. 2005. Framing public discussion of gay civil unions. *Public Opinion Quarterly*.

[48] Recasens, M. et al. 2013. Linguistic Models for Analyzing and Detecting Biased Language. *Proceedings of the 51st Annual Meeting on Association for Computational Linguistics*. (2013), 1650–1659.

[49] Rugg, D. 1941. Experiments in Wording Questions: II. *Public Opinion Quarterly*.

(1941). DOI:https://doi.org/10.1086/265467.

[50] Schuldt, J.P. et al. 2011. “Global warming” or “climate change”? Whether the planet is warming depends on question wording. *Public Opinion Quarterly*. (2011). DOI:https://doi.org/10.1093/poq/nfq073.

[51] Smith, A. et al. 2014. Hiérarchie: Interactive Visualization for Hierarchical Topic Models. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. (2014), 71–78.

[52] Smith, M.S. et al. 1967. The General Inquirer: A Computer Approach to Content Analysis. *American Sociological Review*.

[53] Speer, R. and Havasi, C. 2012. Representing General Relational Knowledge in ConceptNet 5. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. (2012).

[54] Strapparava, C. and Mihalcea, R. 2007. Semeval-2007 task 14: Affective text. *Proceedings of the 4th International Workshop on Semantic Evaluations* (2007), 70–74.

[55] Tausczik, Y.R. and Pennebaker, J.W. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*. 29, 1 (2010), 24–54. DOI:https://doi.org/10.1177/0261927X09351676.