

# An Integrated Approach to Detect Media Bias in German News Articles

Timo Spinde  
Timo.Spinde@uni-konstanz.de  
Universities of Konstanz and  
Wuppertal

Felix Hamborg  
Felix.Hamborg@uni-konstanz.de  
University of Konstanz

Bela Gipp  
Gipp@uni-wuppertal.de  
Universities of Konstanz and  
Wuppertal

## ABSTRACT

Media bias may often affect individuals' opinions on reported topics. Many existing methods that aim to identify such bias forms employ individual, specialized techniques and focus only on English texts. We propose to combine the state-of-the-art in order to further improve the performance in bias identification. Our prototype consists of three analysis components to identify media bias words in German news articles. We use an IDF-based component, a component utilizing a topic-dependent bias dictionary created using word embeddings, and an extensive dictionary of German emotional terms compiled from multiple sources. Finally, we discuss two not yet implemented analysis components that use machine learning and network analysis to identify media bias. All dictionary-based analysis components are experimentally extended with the use of general word embeddings. We also show the results of a user study.

## ACM Reference Format:

Timo Spinde, Felix Hamborg, and Bela Gipp. 2020. An Integrated Approach to Detect Media Bias in German News Articles. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20)*, August 1–5, 2020, Virtual Event, China. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3383583.3398585>

## 1 INTRODUCTION AND RELATED WORK

Media bias, i.e., slanted news coverage, can change the public opinion on any topic heavily [1]. Many approaches to identify such bias exist, however, no automated methods aiming to identify bias in German news texts are available. The objective of this work is to propose, implement and evaluate a system capable of detecting bias words in German news articles. The key contribution of this poster is our media bias identification approach, which includes five components: (i) An IDF-based component, which utilizes word frequencies over a set of documents. (ii) A sentiment-based component using multiple dictionaries. (iii) A component that uses a dictionary of bias words based on semantic models. (iv) A component that uses SVM with cues of historical linguistic development. (v) A network analysis component. Moreover, we provide a summary of characteristics of sentiment in German language.

A helpful technique to find bias are word embeddings, which can be used to find semantically similar words for any given word

[2]. Hube et al. used linguistics and embeddings to address biased language statements in Wikipedia articles [4].

The main shortcomings of prior work are a dependency on manually created resources, a small number of polarity categories and a focus on only specific topics. First, some of these methods identify media bias using predefined dictionaries, requiring manual and effortful creation and adaptation. Second, the possible emotional influence of the detected bias words has not been analyzed on a computational scale. Third, limited research has been conducted on the combination of existing approaches.

## 2 METHODOLOGY AND RESULTS

The methodology proposed in this poster consists of different steps, which are depicted as colored boxes in Figure 1. To train our word embedding model, we used news articles from three national news outlets, *SZ* (65,000 articles), *TAZ* (600,000), and *Südkurier* (286,700).

The automated analysis workflow consists of five components, of which the following three are implemented in the current prototype: an IDF-based component, a combined dictionary-based component, and a component based on semantically created bias dictionary. One of the two components that are not yet implemented will use SVM to analyze historic linguistic cues , such as the pejorative sentiment of the suffix *-ling* in words like *Flüchtling* or *Schönling* [3]. The second component that is not yet implemented relies on network analysis. A variety of nodes, edges, and attributes come to mind, such as newspapers, authors, or bias words. With a sufficiently large data set and further reliable methodology to detect the actual values, topic- and context-dependent patterns could be modeled.

The first component uses IDF scores to measure whether a term is common or rare across the corpus. This way, we aim to find rare words in the collection of articles which are reporting on the same event. Lim et al. propose that, for such a set of news, words with high IDF scores are most likely to be biased words [6]. IDF scores were first calculated among the whole set of articles to be analyzed. We clustered the documents into the even more similar ones by using affinity propagation, and analyzed again.

The second component uses linguistic cues and sentiment to identify bias. The categories in the first version of this dictionary are factive, assertive verbs, entailments, hedges, subjective intensifiers, and one-sided terms [4]. As a foundation of the dictionary that this component is based on, we used the German version of the Linguistic Inquiry and Word Count (LIWC) dictionary [8]. As especially slang and sociolect words are excluded, a separate dictionary by El-Assady et al. [5] was added. In a final step, the dictionary was also extended by assertive verbs, scraped from the *Online-Wortschatz-Informationssystem Deutsch* (OWID) [7]. With these resources, words were classified as bias words if they matched with

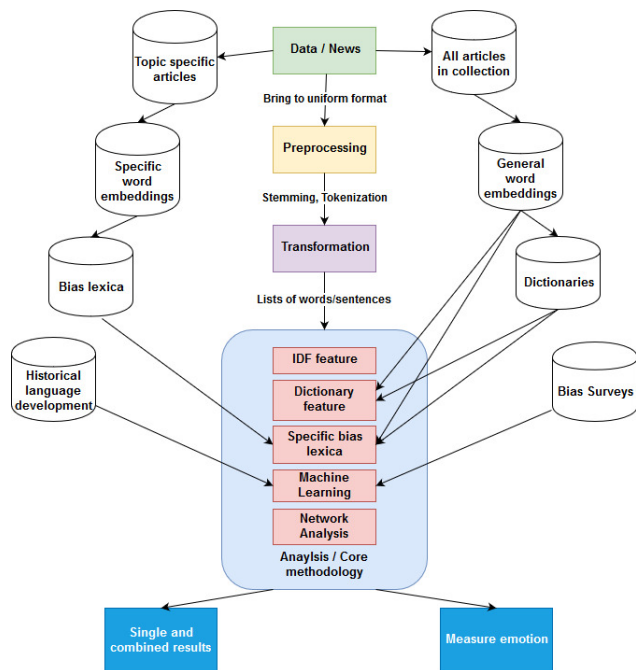


Figure 1: System architecture

any dictionary entry. To improve performance, words were also seen as biased if one of their two most similar terms, as modeled by the word embeddings described before, matched.

The third component uses a topic-specific bias dictionary, based on a separate data set and word embeddings trained on a potentially strongly biased data set. In this case, we used a manually selected set of articles from the newspaper *Bild*. To create a dictionary, seed bias words are manually chose and used to retrieve other bias words. The idea is, as shown in [4], to use word vectors from documents which “are expected to have a high density of bias words.” For each 10-word batch in an initial manually selected list, the 20 most similar words were retrieved and again merged into one list, which hence contains 200 higher potential bias words. This process is then iterated a second time: the 200 words are used as new seed words to extract another 20 most similar words among batches of 10, which leads to an overall of 400 bias words. The full bias dictionary was then added to the dictionary described in the previous section. The overlap was 42%, so most bias terms were not previously included.

To evaluate the approach, we conducted a test, in which we asked 48 participants (mostly students aged between 15-30 years, of balanced gender, from various study programs but without linguistic background, and consuming news daily while not intentionally comparing different media sources) to read three news articles. The same group of articles was shown to 4 persons. For each text, we asked them to highlight bias words, i.e., words they “felt were inducing an assessment.” Only words that were at least mentioned by 2 of the 4 persons in each group were kept. We find that the dictionary component, combined with the topic-dependent bias word dictionary, performed best ( $F_1=0.31$ ,  $P=0.43$ ,  $R=0.26$ ). When considering only adjectives,  $F_1$  was 0.41. Integration of word embeddings did surprisingly not lead to higher accuracy, i.e.,  $F_1 = 0.30$ .

So far, word embeddings added noise, suggesting that our detection might not perform well with synonyms of bias words. This will be one focus of future research. We also aim to improve the overall performance and to create a more reliable evaluation data set, incorporating more participant variables, such as, e.g., political ideology and attitude towards news. Currently, study participants and their results were not sufficiently evaluated.

### 3 CONCLUSION AND OUTLOOK

This poster proposes a work-in-progress approach to identify bias words in German news texts. So far, we implemented three components and tested them in different combinations: an IDF-based component, selecting terms based on their frequency; a dictionary-based component, merging multiple sources of emotional and linguistic terms; and lastly a bias word dictionary that we created using word embeddings. The second and third component combined return the best results with an  $F_1$  score of 0.31 and 0.41 when only considering adjectives. The IDF component and word embeddings (to also identify words similar to our dictionary) do not improve the performance. Using the LIWC in the dictionary-based component, the approach capably identifies emotion markers. Current models and code can be found at <https://zenodo.org/record/3846685#.XsofNsDgqUk>. Even though partial results were promising, the method needs improvement and a more reliable evaluation.

Our approach is a first step towards automatically analyzing bias in German media. Upcoming research will focus on improving the underlying model by enlarging the dictionary, adding more bias dictionaries for individual newspapers, training more reliable word embedding models, gathering a larger amount of data, integrating context and creating a reliable evaluation dataset.

### 4 ACKNOWLEDGMENTS

This work was supported by the Hanns-Seidel-Foundation and the Federal Ministry of Education and Research of Germany.

### REFERENCES

- [1] F. Hamborg, K. Donnay, and B. Gipp. “Automated identification of media bias in news articles: an interdisciplinary literature review”. In: *International Journal on Digital Libraries* (2018), pp. 1–25. doi: 10.1007/s00799-018-0261-y.
- [2] F. Hamborg, A. Zhukova, and B. Gipp. “Automated Identification of Media Bias by Word Choice and Labeling in News Articles”. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, event-place: Urbana-Champaign, IL, USA, June 2019. doi: 10.1109/JCDL.2019.00036.
- [3] R. Harnisch. “Das generische Maskulinum schleicht zurück. Zur pragmatischen Remotivierung eines grammatischen Markers”. In: *Formen und Funktionen. Morphosemantik und grammatische Konstruktion* (2016), pp. 159–174.
- [4] C. Hube and B. Fetahu. “Detecting Biased Statements in Wikipedia”. In: *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018, pp. 1779–1786. doi: 10.1145/3184558.3191640.
- [5] M. Jawale et al. “Interactive visual analysis of transcribed multi-party discourse”. In: *ACL 2017: the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 49–54. doi: 10.18653/v1/P17-4009.
- [6] S. Lim, A. Jatowt, and M. Yoshikawa. “Towards Bias Inducing Word Detection by Linguistic Cue Analysis in News”. In: *DEIM Forum 2018 C1-3* (2018).
- [7] C. Müller-Spitzer. “OWID-A dictionary net for corpus-based lexicography of contemporary German”. In: (2010).
- [8] M. Wolf et al. “Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count”. In: *Diagnostica* 54.2 (2008), pp. 85–98. doi: 10.1026/0012-1924.54.2.85.

**Listing 1: Use the following BibTeX code to cite this article**

```
@InProceedings{Spinde2020a,  
  Title = {An {Integrated} {Approach} to {Detect} {Media} {Bias} in {German} {News} {  
    Articles}},  
  Author = {Spinde, Timo and Hamborg, Felix and Gipp, Bela},  
  Booktitle = {Proceedings of the {ACM}/{IEEE} {Joint} {Conference} on {Digital} {  
    Libraries} ({JCDL})},  
  Year = {2020},  
  Month = {Aug.},  
  Topic = {newsanalysis},  
}
```