The Guidelines Specialists at FIGNEWS 2024 Shared Task: An annotation Guideline to Unravel Bias in News Media Narratives Using a Linguistic Approach

Ghizlane Bourahouat
School of Information Science,
Rabat, Morocco
ghizlane.bourahouat@esi.ac.ma

Samar M. Amer
Freelance Linguist,
Egypt
samaramer.linguist@gmail.com

Abstract

This article presents the participation of "The Guideline Specialists" in the FIGNEWS 2024 Shared Task, which aims to unravel bias and propaganda in news media narratives surrounding the Gaza-Israel 2023-2024 war. Leveraging innovative annotation methodologies and drawing on a diverse team of annotators, our approach focuses on meticulously annotating news articles using a linguistic approach to uncover the intricate nuances of bias. By incorporating detailed examples and drawing on related work that show how language structure represented in the use of passive voice or the use of nominalization and the choice of vocabulary carry bias, our findings provide valuable insights into the representation of the Gaza-Israel conflict across various languages and cultures. The guideline we developed detected the bias against Gaza, against Israel and others by setting keywords that are based on linguistic background tested by the AntConc concordance tool. The result was an annotation guideline that have a solid base. Through this collaborative effort, we developed a guideline that contributes to fostering a deeper understanding of media narratives during one of the most critical moments in recent history.

1 Introduction

The Gaza-Israel conflict of 2023-2024 has not only been a pivotal moment in recent history but has also underscored the critical role of media narratives in shaping public perception and understanding of complex geopolitical events. In response to the multifaceted portrayal of this conflict across various news outlets, the FIGNEWS 2024 Shared Task, entitled "Framing the Israel War on Gaza: A Shared Task on News Media Narratives," emerges as a timely and imperative initiative. This shared task, as outlined in the FIGNEWS Overview (Wajdi Zaghouani et al., 2024), seeks to study and examine the biases and double standards prevalent in news media narratives surrounding the Gaza-Israel conflict.

Adhering to the format prescribed by the shared task's title, the FIGNEWS Overview paper meticulously outlines the objectives and methodology of the initiative, emphasizing the need for a comprehensive exploration of media narratives to foster a nuanced understanding of the conflict. This alignment underscores the shared task's commitment to addressing the overarching goal of unravelling bias and propaganda within news articles, thereby contributing to a more informed and balanced discourse on the Gaza-Israel conflict.

In recent years, shared tasks and similar collaborative efforts have proliferated within the Natural Language Processing (NLP) community, reflecting a growing recognition of the importance of large-scale datasets and standardized annotation guidelines in advancing research in this field. Works such as the SemEval series (Nabil Hossain et al., 2020) and the CoNLL Shared Tasks (CoNLL, 2019) have played pivotal roles in facilitating benchmarking, fostering innovation, and driving progress in NLP tasks, including sentiment analysis, named entity recognition, and discourse parsing.

However, while these initiatives have significantly contributed to the advancement of NLP research, they often focus on generic linguistic tasks and lack the specificity required to address domain-specific challenges such as media bias analysis. In contrast, the FIGNEWS 2024 Shared Task represents a novel and targeted effort to tackle the complexities of bias and double standards in news media narratives, particularly in the context of a high-stakes geopolitical conflict.

Proceedings of The Second Arabic Natural Language Processing Conference, pages 672–676

By honing in on this specific domain, the shared task not only offers a unique opportunity to deepen our understanding of media representations but also paves the way for developing specialised methodologies and techniques tailored to address these challenges.

In this introduction, we situate the participation of our team, "The Guideline Specialists," within the broader context of shared tasks in the NLP community, highlighting the distinctiveness of the FIGNEWS 2024 Shared Task and the novel contributions it offers to the field. Through our engagement in this initiative, we aim to advance the state-of-the-art in media bias analysis and contribute to a more nuanced and informed discourse on the Israeli war on Gaza by introducing a linguistically and statistically informed guideline to guide the annotation of the news under investigation. The next section presents the methodology that is used to develop our guideline to detect bias in the media news on Israel's war on Gaza.

2 Annotation Methodology and Examples

The guideline followed two approaches. The first approach is a linguistic one that is partly based on the work of (Söğüt, 2018) that represents how bias is detected through language structure, especially in the use of the passive voice. Verbs in the passive voice of violent and destructive meaning that are not followed by any reference to the doer are biased against the recipient/victim of these violent actions. “The passive voice can also be used to purposely leave the performer of the action unspecified, a strategy that the reporter can use to avoid ascribing direct responsibility for an action to anybody in particular” (Busa, 2014, 102). The second part of the linguistic approach according to (Söğüt, 2018) is the use of nominalization to avoid mentioning any doer of the actions making the action with no doer that receives no accountability which is biased against the receiver of this action. Nominalization is the “transformation which reduces a whole clause to its nucleus, the verb, and turns that into a noun” (Fowler, Hodge, Kress &Trew, 1979). The third linguistic criterion is lexicon-based as the choice of certain words that label certain groups to direct the reader not to sympathize with them. On the other hand, using words referring to relatives, friends and different relationships in the narrative draws on moving the feelings of reader to sympathise with the group with these more words that refer to family members and to feel angry with the other party which represents bias against this party in media narratives. Examples of the different methods will be presented in the following sub-sections.

2.1 Development of Annotation Guidelines

Seven labels of the Annotation guideline were specified by the shared task description, yet the guideline must put the rules and define the border of each of these labels. Our team’s challenge was to determine the criteria of each label and the language segments to be present in the piece of news to attach a certain label. Each of the seven categories that will be used has some criteria to apply to the news. We have set these criteria after using AntConc and LancsBox to see the frequency of some tokens, the likelihood of the collocation and the concordance in context. Here are the guidelines for each category. The criteria related to the structure level were easy to set. However, those related to the lexicon must be tested by the AntConc concordance tool which was more helpful than the collocation measure as the concordance tool is more comprehensive, and its range can be

Labels	Criteria
Unbiased	News that uses words that state what happened concerning **both sides** with minimum use of adjectives and emotions.
Biased against Palestine	News containing passive voice verbs that refer to killing or destruction with no mention of the doer: For example: “were killed” Use of nominalization to hide the doer: words of bombing and destroying without mentioning the doer, e.g., “bombardment”, “destruction” Examples on the word level are “brutal” “extremists”, “militant”, “Hamas” Examples of words producing anger and bias against Palestine are “mother”, “son”, and “husband” + Israeli names.
Biased against Israel	News containing Words of accusation of aggression violation and genocide, e.g. “genocide”, “violate + convention and “oppression”

Biased against both Palestine and Israel	News containing blame on both parties or accusing both of doing violence.
Biased against others	News containing names of entities as parties to blame or accuse of (expected) violent or complicity with some party and words denoting rebellions and militias, e.g., "Iran", "Houthi", "Hezbollah", "Russia", "rebel", UNRWA", South Africa
Unclear	News context that does not show any direct reference to any party to be responsible of violent actions
Not Applicable	News that does not state information that is relevant to Israel war on Gaza.

Table 1: Annotation labels description.

adjusted to fit very long pieces of news. The seven labels were defined as follows:

2.2 Data Annotation Process

The annotation is done manually using the guideline examples. Detecting the structures and the example keywords in the different pieces of news was the main objective to give the right label to the right news. Ambiguity is treated by detecting frequencies in the piece of news and if a keyword is repeated in the same piece, then the news falls within the label of the keywords. The keywords for the "Biased against others" label always have priority; if the news is ambiguous and the "Biased against others" keywords appear, the choice will be in favour of this label.

2.3 Inter-Annotator Agreement (IAA) Analysis

In conducting the Inter-Annotator Agreement (IAA) analysis for the FIGNEWS Shared Task on News Media Narratives, a robust methodology was employed to ensure the reliability and validity of the results. The IAA methodology involved several key steps, beginning with the selection of a representative sample of data for annotation. This sample encompassed a diverse range of news media narratives, capturing various linguistic structures and thematic elements.

Annotators were provided with detailed annotation guidelines and examples to facilitate a clear understanding of the task requirements and criteria for annotation. These examples played a pivotal role in achieving reliable agreement by illustrating typical instances of annotation and highlighting potential areas of ambiguity or interpretation (Jennifer D'Souza & Sören Auer, 2021).

Once the annotation process was completed independently by each annotator, the annotated data sets were compared, and inter-annotator agreement measures, such as Cohen's Kappa or Fleiss' Kappa, were calculated. These measures quantified the level of agreement between annotators beyond chance, providing insights into the consistency and reliability of the annotation process.

Additionally, discrepancies between annotations were systematically analyzed, with particular attention paid to cases where agreement was low. This analysis identified areas of ambiguity or disagreement, allowing for the refinement of annotation guidelines and the enhancement of annotator training to improve future agreement levels.

Overall, the IAA analysis methodology employed rigorous procedures to ensure the reliability and validity of the agreement measures obtained, enhancing the quality of the annotated data sets for the FIGNEWS Shared Task.

3 Team Composition and Training

The team composition and training for the FIGNEWS Shared Task on News Media Narratives were carefully structured to leverage diverse expertise and ensure effective collaboration. Annotator demographics were a key consideration, reflecting a commitment to inclusivity and cultural sensitivity. The team consisted of two female members: Samar Amer, a linguist from Egypt, and Ghizlane Bourahouat, a data scientist from Morocco. This diverse composition enriched the annotation process by providing insights into linguistic nuances and cultural contexts specific to Egypt and Morocco.

Training sessions were conducted with precision to equip team members with the requisite skills and knowledge for accurate annotation. These sessions covered a spectrum of topics, including annotation guidelines, data interpretation techniques, and quality assurance protocols. Samar and Ghizlane,

drawing on their respective expertise, played integral roles in facilitating the training process, offering valuable insights and guidance to ensure consistency and precision in annotation practices.

Effective team coordination was paramount in optimising workflow efficiency and maintaining annotation quality. Regular meetings and open communication channels facilitated seamless collaboration among team members, allowing for the timely resolution of challenges and the exchange of insights.

4 Task Participation and Results

The Guidelines Specialists' participation in the FIGNEWS Shared Task on News Media Narratives was marked by a meticulous and systematic approach, which yielded highly reliable results. Central to their methodology was the strategic use of detailed examples and comparative analysis with related work, both of which significantly informed their annotation process and outcomes.

The use of examples played a pivotal role in guiding the annotation process. Detailed examples were included in the annotation guidelines to serve as clear benchmarks, helping to minimize ambiguities and ensure consistent application of the guidelines across various pieces of news. These examples were particularly beneficial during training sessions, where they provided practical references that enabled annotators to internalize the guidelines effectively and apply them uniformly.

In addition to leveraging examples, the team conducted a comprehensive review of related work in the field of news media narrative analysis. By comparing their methodology with established practices and findings from previous studies, the team was able to adopt proven techniques and avoid common pitfalls. This comparative approach informed their strategy, leading to more accurate and consistent annotations.

We annotated 1509 sentences. Descriptive analysis of our annotations shows that the 'biased against Palestine'label is the most common type in the annotated corpora that was equally distributed as 25% (Arabic), 25%, (English), 25% (French), 25% (Hindi) and 25% (Hebrew).

The effectiveness of our team's approach was validated through the calculation of Cohen's Kappa, a measure of inter-annotator agreement. The team achieved a Cohen's Kappa score of 85%, as depicted in Figure 1, indicating a high level of agreement among annotators. This high Kappa score reflects the annotation guidelines' clarity and the training sessions' success in preparing the annotators.

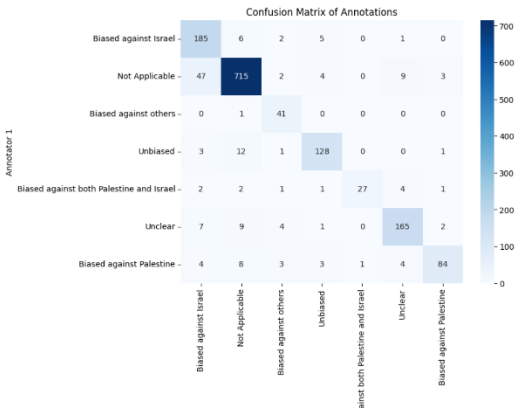

Figure 1: IAA Correlation Matrix**.**

The annotation scores for "The Guideline Specialists" on project B01 using the SQuAD dataset, as evaluated by the organization team, are as follows: a precision of 0.003, a recall of 0.098, and an F1 score of 0.068. The overall performance score for the team is 0.740.

5 Discussion

The team's findings from the FIGNEWS Shared Task on News Media Narratives offer significant insights into the detection and analysis of bias in news media. By employing a well-defined methodology that incorporated detailed examples and comparative analysis with related work, the team not only achieved high reliability in their annotations but also contributed valuable knowledge to the field of media studies and computational linguistics.

The high Cohen's Kappa score of 0,85 underscores the effectiveness of the annotation guidelines and the training process. This level of agreement among annotators highlights the clarity and precision of the guidelines, which were informed by previous work, particularly the studies of (Söğüt, 2018) and (Busa, 2014, 102) on linguistic indicators of bias such as passive voice and nominalization. The use of these linguistic features as part of the annotation criteria proved effective in identifying bias, aligning with the findings of related research.

Our team's approach and findings also have broader implications for the study of media narratives. By systematically applying linguistic criteria to detect bias, the team demonstrated how

structured and replicable methods can enhance the objectivity and consistency of media analysis. This methodological rigour contributes to the ongoing efforts to develop automated tools for bias detection, which are crucial in an era of increasing media consumption and misinformation.

Furthermore, the team's work enriches the discourse on how specific linguistic structures and word choices can influence readers' perceptions. The findings support the notion that passive structures and nominalizations can obscure responsibility and accountability, thereby introducing bias. This aligns with Fowler et al.'s (1979) observations on the manipulative potential of language in media.

In comparison with related work, such as the work of (Md. Adnanul Islam et al., 2022) that presented comprehensive guidelines for consistent and effective annotation of text corpora with emotion labels that led to substantially (30%) higher agreement scores among human annotators, the team's focus on combining linguistic criteria with practical annotation tools like AntConc and LancsBox showcases an innovative approach to bias detection. This integration of computational tools with linguistic theory not only enhances the accuracy of bias identification but also provides a replicable framework that can be applied to other datasets and languages.

The annotation scores for the team "The Guideline Specialists" on project B01 using the SQuAD dataset, as evaluated by the organization team, demonstrate performance with notable variance. The key metrics reveal a precision of 0.003, a recall of 0.098, and an F1 score of 0.068. These results indicate that while the recall is relatively higher, suggesting the team captures a fair number of relevant instances, the precision and F1 score is quite low, pointing to challenges in achieving accuracy and overall balanced performance. With an overall performance score of 0.740, there's room for improvement, particularly in enhancing precision to improve the F1 score.

Overall, the Guidelines Specialists team's findings contribute to the field by providing a validated and replicable methodology for bias detection in news media narratives. The high inter-annotator agreement achieved demonstrates the potential for these methods to be used in larger-scale studies, potentially informing the development of automated systems for real-time bias detection in media. This work advances academic understanding and practical applications in promoting media literacy and accountability.

6 Conclusion

The FIGNEWS Shared Task on News Media Narratives demonstrated the critical role of thorough examples and a deep understanding of related work in achieving high-quality annotations. The use of detailed examples ensured consistent application of guidelines, while insights from related studies informed a robust methodological framework. The team's success, reflected in the high inter-annotator agreement score, highlights the effectiveness of combining linguistic theory with practical annotation tools. These contributions not only advanced the task but also provided valuable methodologies for future research in media bias detection.

References

Busa, G.M. (2014). *Introducing the language of the news: A student's guide*. Routledge.

Fowler, R., Hodge, B., Kress, G. and Trew, T. (1979). *Language and social control.* London: Routledge.

Jennifer D'Souza & Sören Auer. (2021). Sentence, phrase, and triple annotations to build a knowledge graph of natural language processing contributions—A trial dataset. Journal of Data and Information Science, 6(3).

Md. Adnanul Islam, Md. Saddam Hossain Mukta, Patrick Olivier, & Md. Mahbubur Rahman. (2022). Comprehensive guidelines for emotion annotation. Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents, 1–8.

Nabil Hossain, John Krumm, Michael Gamon, & Henry Kautz. (2020). SemEval-2020 Task 7: Assessing Humor in Edited News Headlines. In the Proceedings of the Fourteenth Workshop on Semantic Evaluation.

Söğüt, S. (2018). Ideology in the news through active, passive sentences and nominalization: A study on the terrorist attack in Ankara reported in British and American Newspapers. Journal of Language and Linguistic Studies, 14(1), 162-177.

Wajdi Zaghouani, Mustafa Jarrar, Nizar Habash, Houda Bouamor, Imed Zitouni, Mona Diab, Samhaa R. El-Beltagy and Muhammed AbuOdeh . 2024. The FIGNEWS Shared Task on News Media Narratives, In Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024). Association for Computational Linguistics, Bangkok, Thailand.