Exploring Gender Bias in Political News Articles from African Countries

Nina ten Pas 7715234
Sterre Rood 6545475
Julia Ruiter 6689922
Eveline Schmidt 9975055

1. Abstract

Much research has been done on the representation of women in political news. However, most studies focus on one specific country. The purpose of this study is to examine whether there is a difference in gender bias in political news articles between African countries and if this correlates with a different gender makeup of government. To answer the research question, abstracts of 98000 English news articles from 4 Nigerian and 3 South African sources are used. Using various NLP techniques, we uncover inconclusive evidence and contradictory correlations.

2. Introduction and Background Information

Generally speaking, women are underrepresented in political participation worldwide. Even though the proportion of women in national parliaments around the world is growing, they are still underrepresented. These days, gender inequality in the amount of news media coverage politicians receive is an important subject as contemporary politics is highly mediatized, and voters rely almost exclusively on news media as their source of political information. Unfavourable news media coverage can harm the electoral chances of politicians and threaten the political longevity of sitting politicians [1].

Studies have shown that news coverage of politics often reflects gender bias, with women having a harder time making the news than men [2], [3]. For instance, the study of O'Neill and Savigny [4] showed that women have a harder time making the news in the British political press. Furthermore, Dacon and Liu [5] found that women are mostly associated with family-related words, such as: wife, mother, family and home. Whereas men are mostly associated with political and business-related words, like: president, manager, mayor and economy. The way these words are used in news media massively affects gender roles. Another study of Marjanovic, Stanczak and Isabelle Augenstein [6] found that in comments on Reddit about politicians, there was an equal public interest in both female and male politicians. This might suggest that there is a positive change happening regarding gender roles in politics. However, the opposite seems to be true, as the interest in female politicians was based on completely different grounds. Female politicians were addressed less respectfully than their male counterparts. The Reddit users were more focussed on the appearance and family of the female politicians, whereas for the male politicians this was not the case. Another interesting difference was the way the politicians were addressed. The females were mostly addressed by their first name, unlike the males. This disrespectfulness was most present in extremely right oriented discussion groups. It was, however, also found in other less extreme right- and left-oriented discussion groups, which shows that this problem is not only happening in the extreme groups in our society.

The way female politicians are appearing in the media discourages women from participating in politics. Women are not given an equal voice. Since the news media has the power to shape the public perception and political chances it is important to research gender representations in the political news media [4].

A lot of research has been done on the representation of women in the news media in general. Over the past few years, bias in news articles, news images and news videos has been studied [7], [8]. Most studies find that there is an unequal representation of women in the news, compared to men. For example, the research from Docan and Liu [5] found that females suffer from socially constructed biases in the news. The research from Rao and Taboada [9] found an unequal gender representation in certain topics in news articles. This underrepresentation of women is present in various areas of the news [10], [11].

Focusing on research about the representation of women in political news, much research has already been done. The research from Hooghe et al. [12] focused on Belgian politicians. They found that even though the Belgian political system evolves rapidly towards equal representation, gender bias in the media is persistent. The study from Ette [3] focusses on the representation of Nigerian women politicians in the news media. Even though Nigerian women have become more active in politics, the study found that engagement in politics is not reflected in media coverage. The research from Thomas et al. [13] focusses on female heads of government in Canada. They show that fewer stories about female heads of government are written compared to male heads of government. Furthermore, they conclude that news about heads of government remains gendered.

Thus, much diverse research has been done on how women are represented in the news media. This also goes for research on the representation of women in political news. However, most studies focus on one specific country. This study contributes to current scientific research by investigating the representation of women in political news articles compared across African countries. Considering the significant difference between the African countries in terms of the number of women in politics, this comparison may be of interest and may contribute to the current scientific research.

In conclusion, the previous studies show that there are two gendered patterns [14]: female politicians are underrepresented as in how often women are mentioned in political news compared to male politicians [4], [15] and women are represented in a stereotypical way as stated in the study of Marjanovic, Stanczak and Isabelle Augnestein [6]. This study focuses specifically on gender bias in terms of the extent to which women are underrepresented in political news media. Thus, the focus is on the first gendered pattern: how often female politicians are mentioned in political news articles compared to male politicians. As indicated above, this gender bias is examined between African countries. This leads us to our research question: *is there a difference in textual features in political news articles which may indicate a gender bias difference between African countries?*

This study specifically looks at political news articles from the countries Nigeria and South Africa. This comparison is interesting because Nigeria has 5.6% female representation in politics and South Africa has 42.7% female representation in politics [15]. This difference in female representation in politics may indicate that there is a difference in gender bias in political news articles in both countries. As the focus of this study will be solely on textual features and not combined with videos or images, this research is monomodal. Looking at previous research, which showed that there is a prevailing gender bias toward males in news in general, the expectation is that in both countries political news articles are biased toward males. Thus, our first hypothesis (H1) is: *there is a gender bias towards males in political news articles from South Africa and Nigeria*. However, given the significant difference in female representation in politics, this bias towards males is expected to be less strong in South African political news articles than in Nigerian political news articles. Thus, our second hypothesis (H2) is: *the gender bias towards males is less strong in South African political news articles compared to Nigerian political news articles.*

3. Data

Raw Data

The corpus used for this research consists of roughly 98000 article abstracts from 3 South African news sources (news24.com, iol.co.za and ewn.co.za) and 4 Nigerian news sources (punchng.com, dailypost.ng, pulse.ng and premiumtimesng.com). Since the focus of this research is on differences in gender biases among the political domain, the first objective of the data preparation process is to retrieve all political articles of these abstracts. The article abstracts were made available by a correspondent at Owlin B.V., a news and text- analytics company based in Amsterdam.

Preparing the Data

Retrieving the political articles out of all abstracts can be turned into a classification problem. Classification in machine learning means trying to identify which set a new observation belongs to [16]. One of the most popular tools for binary textual classification is Logistic Regression due to its accessible interpretation. As input data, Logistic regression requires a numerical representation of the words and a labelled dataset to train the classifier on.

To represent the words as numerical features, TF-IDF is used to vectorize the words into numerical vectors. It does this by counting how statistically relevant a word in a certain document is compared to the other words and documents [17]. This way, TF-IDF highlights words that are meaningful to an abstract and may distinguish a labelled abstract from one labelled differently.

The labelled abstracts are fetched by manually labelling 300 abstracts as ‘political’ or ‘not political’. After using the TF-IDF Vectorizer of scikit learn on these abstracts, the Logistic Regression model is trained using these vectors. This model is then used for classifying 2000 articles as ‘political’ or ‘not political’ to keep the results interpretable. These results are overviewed manually again where 6% of the results are manually changed into the right classification. Finally, these 2300 labelled abstracts and their TF-IDF vectors were used to train a new Logistic Regression, which was used to classify all 98000 articles as ‘political’ or not. As a result the political articles found for South Africa and Nigeria were 7176 and 6362, randomly balanced down to two equal sets of 6362 political abstracts per country, resulting in the final dataset used for this research.

Descriptive Statistics

For the South African political abstracts the mean length of the abstracts is 613 words (standard deviation = 1167). But the high variance shown in **Figure 1** indicates the lengths can vary a lot, which as well can be seen when looking at the maximum and the minimum lengths of the abstracts. The maximum length of an abstract was 7000 words for the South African news articles, whereas the minimum was 37 words. This difference in length is due to the way the articles were scraped from the news sources. There were similar results found for the Nigerian political abstracts lengths differences, where the maximum length was 9969 words and the minimum only 5 words. The mean length of the Nigerian abstracts, however, was 394 words (standard deviation = 308), which is much smaller than that of South Africa. The widespread of the data is also shown in the large variance of both countries, being 1,361,583 for South Africa and 94,762 for Nigeria.

```
South African abstracts:
variance: 1361583.2948362685
std: 1166.869013572761
mean: 613.162100456621
max: 7000
min: 37

Nigerian abstracts:
variance: 94762.36174887787
std: 307.8349586204885
mean: 393.6391022744389
max: 9969
min: 5
```

Figure 1 Abstract Statistics

To create a better understanding around the interpretations of the abstracts used, the most frequently used nouns per country are visualized in **Figures 2 and 3**. Both word clouds show that the abstracts are indeed very politically oriented as can be seen by words such as 'government', 'election' and 'state'. The words do not seem to show any teasers about what to expect related to gender biases. However, this could be due to the fact that English is not a grammatical gender language.

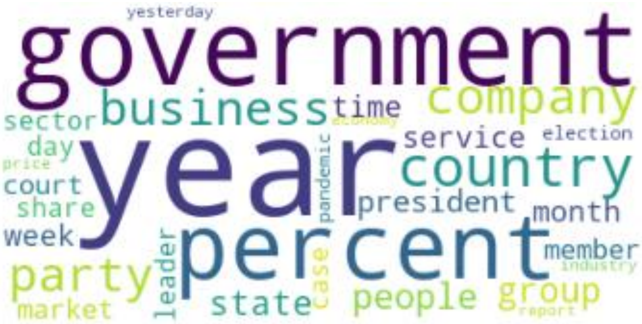

Figure 2 Word cloud of most frequently used nouns in South African abstracts

Figure 3 Word cloud of most frequently used nouns in Nigerian abstracts

Quality of the Data

Regarding the quality of the dataset, it is important to mention that the way the abstracts were scraped differs due to the varying website constructions. As a result, some abstracts were cut off in the middle, therefore parts of the abstracts could be missing. Examples can be found in **Figures 4 and 5**. Missing parts of the abstracts could influence the results of the analysis on the abstracts, as other abstracts are complete, so inconsistency in the data is found. Another finding that affects the quality of the data, and therefore potentially the results, is the fact that for some news sources the title of the article is repeated in the abstract. An example of this can be seen in **Figure 5**. This could influence the results later on, since, for example, proper nouns could be mentioned twice when the title is repeated.

The Lagos State Governor, Mr Babajide Sanwo-Olu, has called on residents of the state to prioritise the welfare of the elderly population of the state. The governor was quoted to have made the call while inaugurating the Thani-Oladunjoye Old People Care Centre in Oke-Owode, Epe, according to a statement on Tuesday. Sanwo-Olu, who was represented… Read More

Figure 4 Example of an abstract that was cut off

The Director-General of the Progressives Governors Forum (PGF), Dr. Salihu Lukman has accused the Chairman of the Caretaker/Extraordinary Convention Planning Committee (CECPC) of the All Progressives Congress (APC) of manipulating the party's decision as regards the February date fixed for the national convention. He said the party needed leaders that would really be sincere and […] APC Convention: DG Progressive Govs Forum accuses Buni of manipulating party's decision

Figure 5 Example of an abstract that was cut off and where the title has been added

4. Methods and Various Approaches

Methodology

The following section explains the different approaches to finding biases in our political abstracts. As many methods have outcomes that help interpret our following method, and they help by an overall interpretation of the current method, we will shortly touch upon some results. The first half of our method section focuses on using word embeddings to detect biases in our abstracts, whereas our second half focuses on gender distributions provided by word counts.

Approach 1: Word Embeddings – Top 30 Nouns

The first approach of this research is focused on looking at the differences between female and male word embeddings. Word embeddings are distributed representations in a vector space [18]. One of the most frequent techniques to compute such vectors is Word2Vec. In this research we have used the skipgram variant to calculate the embedding tables. Skipgram consists of a single hidden layer neural network to predict the context of a word. After tokenizing, lowercasing and removing punctuations from our abstracts, Word2Vec can be used to calculate the embedding tables of all words. The vectors within these embedding tables consist of vector representations of the words.

The retrieved vectors can be used to calculate the semantic distances between words. Word embeddings have therefore been a useful tool to highlight the semantic differences in social biases like gender bias [19]. To look at the gender bias within our article abstracts, we first needed a female and male representation in the vector space. To calculate these representations, we took the average embeddings of two lists containing female and male terms provided by Dacon et. al [5], but was adapted by our team to remove words that are frequently used as neutral, like “author” and “minister”. After calculating the average embeddings, we need other embeddings from our abstracts to compare the semantic distances to.

The first approach here to calculate the differences in gender per country is to compare the distances of the top 30 most frequently used nouns in the abstracts per country. The distances were calculated with the Euclidean distance, which is a popular technique for evaluating semantic similarity using word embeddings [20]. The nouns are found with Spacy, a library that provides a variety of practical text processing tools [21]. The top 30 are constructed by the most frequently used nouns in our abstracts per country.

Having a list of the nouns per country, we look at the distances of our average male and average female embeddings compared to all words within those noun lists. We then compute the average distance per male (average_distance_male) and female (average_distance_female) for both countries. For each country, we calculate the distance between average male and average female vectors to normalize the relative gender bias by computing the maximum possible gender bias difference. We continue by subtracting the average_distance_female from the average_distance_male and divide this by the normalizing factor for the country. The result is our bias, where:

- A negative score means the average normalized Euclidean distance of females is larger and it is biased towards males. This can be interpreted as the average percent more male bias a country has.
- A positive score means the female distances are smaller and it is biased towards females. This can be interpreted as what percent more female bias is present for each country.
- The closer the score is to zero, the less bias exists.

Figure 6 shows the biases found using the top 30 nouns. There is a notable additional bias towards men in both countries, which is a first indication of the political abstracts being biased towards men overall. A slightly larger male bias outcome for South Africa appeared, but to investigate differences in gender bias within our abstracts further, we will look into more options of using word embeddings to detect biases.

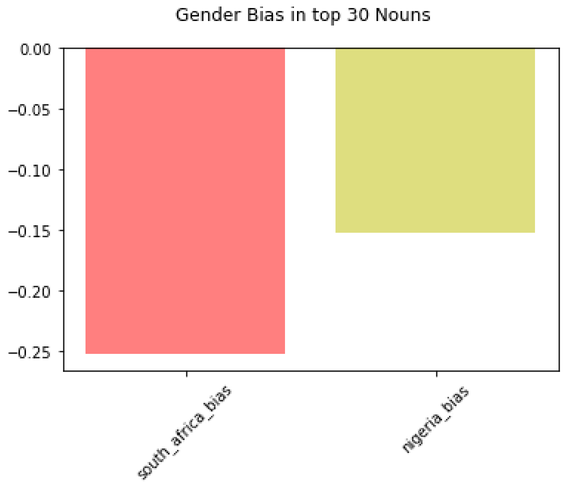

Figure 6

Approach 2: Word Embeddings – Wever's Categories

After receiving scores highly biased towards men using the top 30 nouns, the next approach tries to look deeper into different categories where biases can be found within our abstracts. We have chosen an approach inspired by Wevers et. al [19], that looks into the following (sub)categories: Affective and Emotional Processes; Cognitive Processes; Sensory and Perceptual Processes; Social Processes; Occupation; Leisure activity; Money and Financial Issues; Metaphysical Issues; and Physical states.

All categories contain a list of words relating to the corresponding category, and the same approach as mentioned before, will be used: Using the word embeddings of those words whereafter the Euclidean distances are calculated, and the average male and female distances are calculated which will be used to determine the bias. Looping through all categories and calculating their biases results in the biases seen in **Figure 7**.

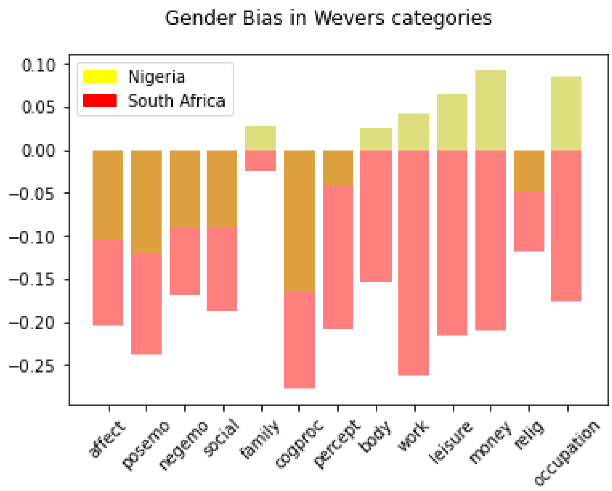

Figure 7

Approach 3: Counter – Female versus Male Articles

Gender bias has been defined above as a difference in representation between genders, but which this means is still up for interpretation. In the earlier sections, we explored the gender bias inherent on the level of the words themselves, but what about the articles as a whole? Difference in representation can also be attributed to a difference in how often a gender is represented in the media. In a perfectly unbiased world we would expect the ratio of political articles per gender to match the distribution of politicians per gender; if a country's article distribution differs from their political gender ratio, then we can say that that country's articles have a gender bias.

To perform this test, a counter method was employed. After lowercasing everything and tokenizing the articles by word, we compare the number of female words to male words in the article abstract, and assign the abstract a gender based on whichever category is larger. The list of gendered words comes from Dacon's study [5]. This simple count of predominantly male versus predominantly female articles was as follows:

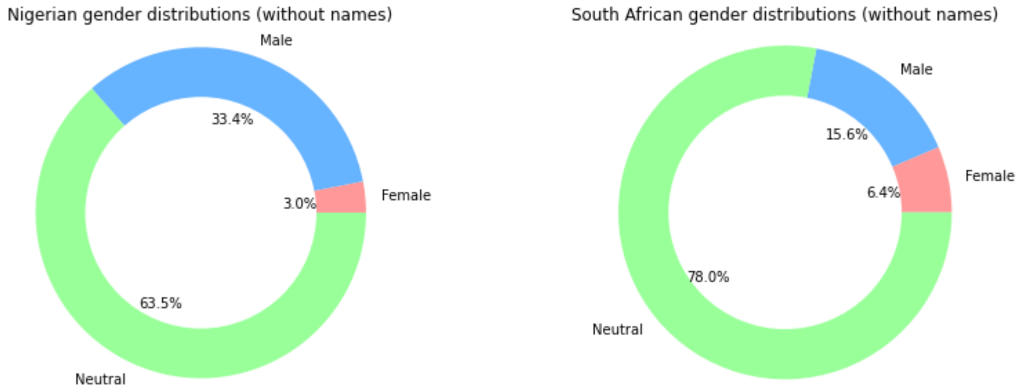

Figure 8

However, we realized that our initial gendered word set did not account for the names of politicians. When investigating the abstracts of the news articles we saw that besides nouns and pronouns that specify a certain gender, there was also a lot of usage of proper nouns. Therefore we have decided to make a list of female and male pronouns for both countries. We did this by investigating the aforementioned sample of politically labelled articles for each country. We went through the abstracts and every time a name was mentioned, the name was put to a list. For each name in the list, it was then investigated what the gender of this person was, resulting in a list of male and female proper nouns per country. As this was a random sample of the whole dataset, we assume that this list of names is a good representation for the whole dataset.

Repeating the counter process on the new set of gendered words (including names) yields the following:

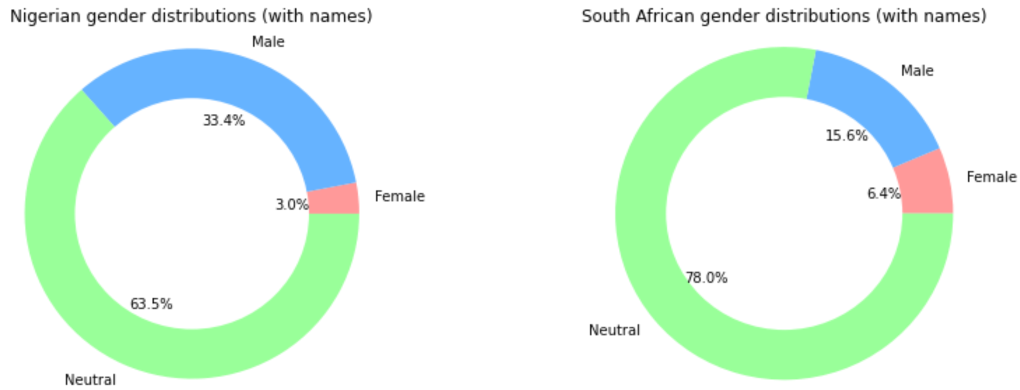

Figure 9

As you can see, the inclusion of names had such a slight impact on article distribution that it didn't change the overall percentages.

In both methods, there are clearly far more “male” articles for both countries. This leads to the conclusion that there is an overall male bias in political news representation. While there is a small case for South Africa being less biased due to the higher percent of neutral articles than gendered articles, comparing the ratio of male to female articles and politicians (using the gender set with names) leads to a different conclusion:

	Percent Male Politicians	Percent Female Politicians	Male to Female Politician Ratio	Percent Male Articles	Percent Female Articles	Male to Female Article Ratio
Nigeria	94.4	5.6	16.857	33.4	3	11.133
South Africa	57.3	42.7	1.342	15.9	6.4	2.484

Figure 10 Political ratio versus article ratio [15]

When comparing the gender ratios of the articles to the gender ratio in politics, we get a counterintuitive result: Nigeria has a lower male to female bias in the articles than in politics, implying that there is actually a ***female*** bias in their media. Meanwhile, South Africa still has an overall male bias.

Approach 4: Counter – Bias within Article Categories

The counter method above highlighted that we had an imbalance dataset, with far more male articles than female articles. To correct this, we decided to analyse the bias per gender category of political news articles. Now, instead of just counting if an abstract is overall male or female (or neutral), we also count the number of male and female words per abstract to get the basis within each category.

For predominantly female articles, this is calculating the female to male ratio (female/male) and number of excess female words (female-male); for predominantly male articles, male to female ratio and excess male words were calculated. By comparing these gender-confirming statistics between the two groups, we can see if there is bias within representations. If the female to male ratio for male articles is lower than the male to female ratio of male articles, then there is a male bias because it may indicate that women are not talked about with regards to themself or their job, but rather talked about by or in regard to their cohorts. Results of this analysis are below:

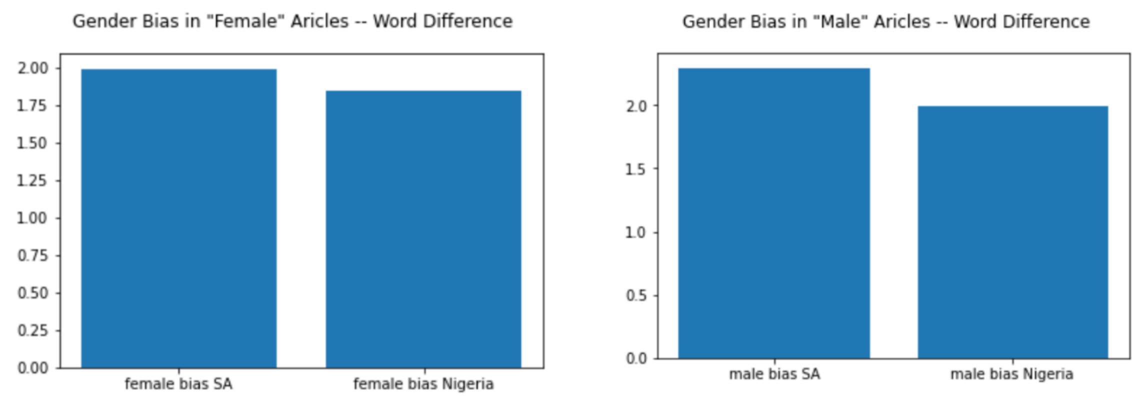

Figure 11

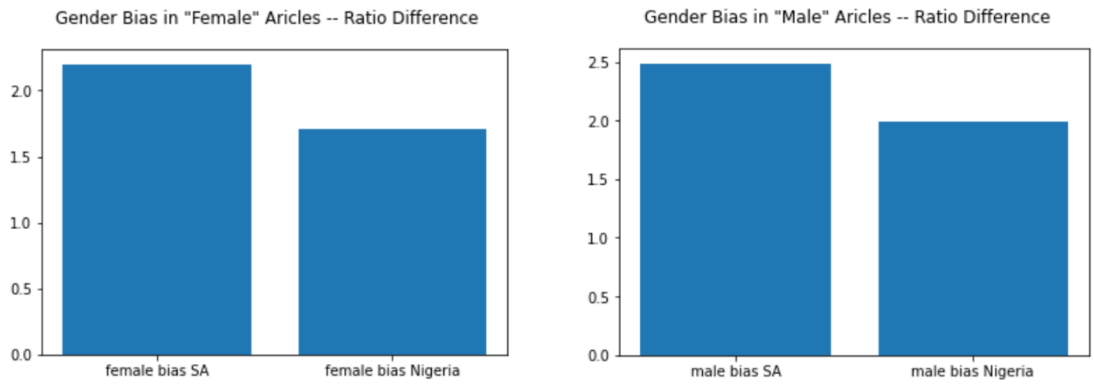

Figure 12

	Word Bias			Ratio Bias		
	Male Bias, Male Articles	Female Bias, Female Articles	Male - Female Difference	Male Bias, Male Articles	Female Bias, Female Articles	Percent more male bias (delta / female bias)
Nigeria	1.992	1.703	**0.289**	1.996	1.842	**8.36%**
South Africa	2.486	2.199	**0.287**	2.295	1.991	**15.27%**

Figure 13 Gender distribution statistics

In both countries, the bias for a particular gender stays strongly in favour of the gender the article is about. For simple differences in excess gendered words, there is neither a noticeable difference between countries nor a sizable bias in gender overall. For ratio comparisons, however, it seems like South Africa has a significantly stronger male bias than Nigeria, once again disproving both of our hypotheses.

5. Results

When comparing the outputs of the counter methods to those of the word embedding methods, it becomes difficult to make a conclusion–the NLP methods yield opposing, nuanced results.
Summary of outcomes:

- **Approach 1:** very large male bias regardless of country, and a meaningful difference between countries – Nigeria averages a bias of around 15% towards men while South Africa averages a bias of around 25% towards men. (**Figure 6**)
- **Approach 2:** bias strength and direction vary by category (individual categories range between 45% more male to 15% more female). Nigeria has 8/13 categories leaning toward a female bias, implying a slight general female bias; every category for South Africa leans heavily male, though. (**Figure 7**)
- **Approach 3:** male bias in terms of representation, but a more female bias in Nigeria when comparing ratio of articles with the ratio of politicians. (**Figures 8, 9, 10**)
- **Approach 4:** slight male bias in terms of gendered word usage within articles as male articles use 0.29 more male words per article than female articles do female words. When comparing gendered word ratios, South Africa has a more prominent male bias (male articles are 15.3% more heavily gendered male than female articles are

female) while the additional male bias for Nigeria is 8.36% more heavily male. (**Figures 11, 12, 13**)

If you only look at the article counters, you would conclude that that political representation and news representation are inverted (resulting in a female bias for Nigeria and a male bias for South Africa); if you only look at the word embeddings methods, you would conclude that political representation has almost no effect on news representation since there is an almost neutral or slight male bias. So what is going on? The output of the first counter method sheds light on this: we have an unbalanced dataset, with far more male articles than female. To counteract this, we needed to consider the gender bias ***per group*** of gendered articles, which was done by adapting the original counter method. But even then, there was still a notable male bias in both countries, with a stronger male bias in South Africa than Nigeria.

6. Conclusion

This study attempted to answer the research question: *is there a difference in textual features in political news articles which may indicate a gender bias difference between African countries?* The textual features we chose to analyse to indicate a gender bias difference between African countries were the nouns, pronouns, and in some cases also the proper nouns, present in political news articles. Previous studies suggest that there is gender bias in political news articles; where most studies looked at one specific (Western) country, this study looked at the difference between two African countries.

In order to answer the research question, two hypotheses were formulated and answers to these hypotheses is provided using the methods explained above. The results show that the first hypothesis '*there is a gender bias towards males in political news articles from South Africa and Nigeria'* is not consistently supported. The findings show that, in most cases, there is a gender bias towards males in political news articles from both countries. However, the strength and the direction of the bias varies by definition of bias and thus the method of analysis.

Looking at both word embedding methods it may be concluded that there is a gender bias in either country. In the South African abstracts, both methods show a bias towards men, whereas for the Nigerian abstracts the methods are not in line with each other. The first approach, word embedding for the top 30 nouns, shows a bias towards males, when in fact the word embedding for the Wevers topic categories show a bias towards females. This is not in line with the first hypothesis. However, it can be argued that looking at these categories, with respect to our political articles, is too detailed to conclude whether there is a general gender bias in political news articles because the categories may not be representative of political media as a whole. As the word embedding for the top 30 nouns reflect the abstracts better than the Wever categories do, it should be concluded that there is a bias towards men for both countries. This bias is quite strong in both cases since most distances are up to 15 - 25 percent more heavily gendered for the target words.

The same inconsistencies in the strength and direction of the biases were also found in the counter methods. With respect to representation in political news articles there seems to exist a male bias for both countries. However, when taking the ratio of articles and ratio of politicians into account, the direction of the bias for Nigerian articles shifts towards females. For South Africa the direction of the bias stays the same.

Given the fact that the first hypothesis cannot be consistently proven, the second hypothesis '*the gender bias towards males is less strong in South African political news articles compared to Nigerian political news articles'* cannot be considered true and is therefore disproven. The only female biases that were found were, counterintuitively, solely in favour of Nigeria. This is against the expectation that South Africa would have a less strong male bias due to its more fair political makeup. Regarding the cases where for both countries a male bias was found, there was again no consistency in the strength of the biases when comparing both countries.

7. Discussion

Limitations

This research has a couple of limitations. Starting with the dataset, the abstracts were cut off in the middle and thus deleting parts of the abstracts. This may have affected the study. For example, if a cut off abstract appears to be neutral from the analyses, it may be that later in the abstract there was more talk about men or women. In addition to the fact that some of the abstracts were cut off in the middle, the abstracts varied significantly in length. For example, a five-word abstract was basically meaningless in this study. This is because a five-word abstract is highly unlikely to contain information about possible gender bias in political news articles. In contrast, some abstracts consisted of thousands of words. This wide variation in the length of the abstracts may have affected the results of the study.

Furthermore, there is also a limitation in the counter method. Given the fact that the abstracts were tokenized and analysed per word, names might be counted double. If a politician was mentioned by first and last name, there will then be two gender points attributed to that article, whereas an article referring only by last name will only get one point. This may affect the overall bias in gendered word counts, but has no effect on the article distribution.

Another limitation that should be mentioned is the fact that only gender specific pronouns and proper nouns were used to determine the genders of the persons named in the article abstracts for the analyses. A lot of politic-specific words are neutral, meaning that they do not show the gender. So if a minister for example was mentioned, not by their name, but as 'Minister of' their field of expertise, such as 'agriculture', this was not counted as a gender specific word even though only one specific person, and thus gender, was meant.

At last, it can be argued that trying out multiple methods introduces bias. When the results were not quite as expected, we did not want to accept that right away and investigated what was going on. In this way we might have brought biases to the research ourselves. The question that has quite often been asked during this study is, when do we stop and accept the results? There are always new things to investigate or to consider. This could be done in future research on this subject.

Future Research

For future research, it may be of interest to be able to conduct analyses on the entire articles, rather than just the abstracts. In this way, no meaningless articles are included, considering the length of the abstracts. This may provide better results given the incomplete abstracts and the wide variation in the length of the abstracts. Thus, examining possible gender bias in (political) news articles can be done more successfully over the entire articles.

It might be a good idea to include a list of ministers along with their field of expertise, in the list of gender words. This should be done per country, as the ministers differ amongst the different countries. It would be interesting to see whether adding these words to the list of gender words will change the results of gender biases for countries with different distributions of males and females within their political parties.

Summary

In summary, this study has shown it is not a simple matter to investigate gender bias in political news articles. The different results from different methods demonstrate this. This also leads to the fact that no uniform answer can be given to the research question. Concluded, "gender bias" is a difficult concept to work with.

8. References

[1] Van der Pas, D. J., & Aaldering, L. (2020). "Gender differences in political media coverage: A meta-analysis". *Journal of Communication, 70*(1), 114-143. https://doi.org/10.1093/joc/jqz046

[2] Tresch, A. (2009). "Politicians in the Media: Determinants of Legislators' Presence and Prominence in Swiss Newspapers". *The International Journal of Press/Politics, 14*(1), 67–90. https://doi.org/10.1177/1940161208323266

[3] Ette, M. (2017). "Where are the women? Evaluating visibility of Nigerian female politicians in news media space". Gender, Place & Culture, 24(10), 1480–1497. https://doi.org/10.1080/0966369x.2017.1387104

[4] O'Neill, D., & Savigny, H. (2014). "Female politicians in the British press: The exception to the 'masculine'norm?". *Journalism Education, 3*(1), 6-27.

[5] Dacon, J., & Liu, H. (2021). "Does Gender Matter in the News? Detecting and Examining Gender Bias in News Articles". *Companion Proceedings of the Web Conference 2021*. https://doi.org/10.1145/3442442.3452325

[6] Marjanovic, Stanczak and Isabelle Augenstein (2021). "Quantifying Gender Biases Towards Politicians on Reddit". *arXiv, 2112.12014.*

[7] Jia, S., Lansdall-Welfare, T., & Cristianini, N. (2015). "Measuring Gender Bias in News Images". *Proceedings of the 24th International Conference on World Wide Web*. https://doi.org/10.1145/2740908.2742007

[8] Peng, Y. (2018). "Same Candidates, Different Faces: Uncovering Media Bias in Visual Portrayals of Presidential Candidates with Computer Vision". *Journal of Communication, 68*(5), 920–941. https://doi.org/10.1093/joc/jqy041

[9] Rao, P. & Taboada, M. (2016) . "Gender Bias in the News: A Scalable Topic Modelling and Visualization Framework". *Frontiers in Artificial Intelligence 4:664737*. https://doi.org/10.3389/frai.2021.664737

[10] Kemble M. (2020). "As Good as the Men? A Corpus Analysis of Evaluation in News Articles about Professional Female Athletes Competing in 'Masculine' Sports". *Crit. Approaches Discourse Anal. Across Disciplines 12* (1), 87–111.

[11] Power K., Rak L., Kim M. (2019). "Women in Business media: A Critical Discourse Analysis of Representations of Women in Forbes, Fortune and Bloomberg BusinessWeek, 2015-2017". *Crit. Approaches Discourse Anal. Across Disciplines 11* (2), 1–26

[12] Hooghe, M., Jacobs, L., & Claes, E. (2015). "Enduring Gender Bias in Reporting on Political Elite Positions". *The International Journal of Press/Politics,* 20(4), 395–414. https://doi.org/10.1177/1940161215596730

[13] Thomas, M., Harell, A., Rijkhoff, S. A., & Gosselin, T. (2020). "Gendered News Coverage and Women as Heads of Government". *Political Communication,* 1–19. https://doi.org/10.1080/10584609.2020.1784326

[14] J. D'Heer, J. Vergotte, S. De Vuyst, and S. Van Leuven, "The bits and bytes of gender bias in online news : a quantitative content analysis of the representation of women in Vice.com," FEMINIST MEDIA STUDIES, vol. 20, no. 2, pp. 256–272, 2020.

[15] “Women in National Parliaments,” Inter-Parliamentary Union: World classification, 01-Feb-2019. [Online]. Available: http://archive.ipu.org/wmn-e/classif.htm. [Accessed: 06-Jan-2022].

[16] Wang, Y., Zhou, Z., Jin, S., Liu, D., & Lu, M. (2017, October). “Comparisons and selections of features and classifiers for short text classification”. *In IOP Conference Series: Materials Science and Engineering* (Vol. 261, No. 1, p. 012018). IOP Publishing.

[17] Qaiser, S., & Ali, R. (2018). “Text mining: use of TF-IDF to examine the relevance of words to documents”. *International Journal of Computer Applications*, *181*(1), 25-29.

[18] Basta, C., Costa-Jussà, M. R., & Casas, N. (2019). “Evaluating the underlying gender bias in contextualized word embeddings”. *arXiv preprint arXiv:1904.08783*.

[19] Wevers, M. (2019). “Using Word Embeddings to Examine Gender Bias in Dutch Newspapers, 1950-1990”. *In Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change (pp. 92-97).*

[20] Omer Levy, Yoav Goldberg, and Ido Dagan. 2015.” Improving distributional similarity with lessons learned from word embeddings”. T*ransactions of Association for Computational Linguistics 3:211–225*.

[21] Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019). “ScispaCy: fast and robust models for biomedical natural language processing”. *arXiv preprint arXiv:1902.07669*.