Oğuz Bulut Kök

Domain Expert: Belgin San-Akca

bakca@ku.edu.tr

**An analysis of the depiction of women in news media based on armed group typology**

**Abstract**

This exploratory study examines how women's depiction in the media differs concerning the armed groups' ideology. For this purpose, news about ethnic and religious groups was collected between the early 1980s and 2010. These pieces of news were divided into topics, and their sentiment polarity, and the frequency of the relation of words "woman" and "female" with other words in the texts were analyzed. The news was usually negative, with religious groups' news being slightly more negative than ethnic news. The news often involved war conditions, civilian life, and international reactions. It is suggested that the women, regardless of the armed groups, are depicted as the victims, often associating them with the hostages and killing them and coupling them with other vulnerable groups. However, the religious and ethnic armed groups differ in the sense that the news about religious groups contains more polarized references to both women as a victim and women as active participants. Further agendas for such research and the evaluation of the tools used were also discussed.

**Introduction**

Many citizens get exposed to some certain cause of a "terrorist" group through the news media. This easily creates some opinions about why the group is composed of terrorists or liberation fighters. Seeing a certain disadvantaged group at peril can affect the attitude of the viewer and their support towards an armed group. In that sense, from news media to recent surge of social media, media representations often influence how audiences perceive the legitimacy, motivations, and actions of armed groups. Therefore, media depictions, especially those whose treatment constitutes norms, very much matter. One of these norms is the treatment of women, which provides a good measure of human rights. However, how can the main motivation for the advent of the group affect these? In other words, how can the group's ideology change how the media refer to these groups? Therefore, to explore this area, the following particular research question was formulated:

How do news portrayals of armed groups differ concerning women in relation to their ideologies?

**Literature Review**

There are few studies, if any, that directly examine the different political backgrounds of the armed groups’ (following the distinction from Lutz & Lutz (2019), ideological, nationalist, and religious groups) depiction when it comes to the role of women and gender in general. Such exploratory research can help to see how different political causes can affect the depiction of women (at least on paper) to certain norms in discourse and how these actors depict these norms to signal their position and garner support in multiple ways. News media serves as the main source of how armed groups expose themselves to the international community. In that sense, the condition of women serves two purposes to better understand these norms within conflict: 1) Women are a disadvantaged group in conflict zones (Cockburn, 2013; Cohen, 2017); therefore, understanding how women are depicted in the media in the news related to armed groups can also help to further understand how and why these groups internalize and support a certain treatment of gender roles. 2) Media depiction can provide details on how these groups are depicted as such and if they differ in relation to their cause. 3) Women's rights might also serve as a starting point to measure human rights' adherence in general since it is a primary component of human rights.

Non-state groups' adoption of human rights and norms has been widely discussed in the literature. The groups often claim to follow human rights as a matter of self-image, aligning with their goal, public relations/recognition, strategic advantage, and acceptance and support in the occupied region (Bangerten, 2011). In addition, one pragmatic component of norm signaling is that these groups often seek reputation and legitimacy for potential support for their cause, often from a state actor (San-Akca, 2016). Gender issues, especially the condition of women, come at the front when it comes to such signaling. For example, a survey experiment conducted by Manekin and Wood (2020) shows increasing support towards an armed group if women are mentioned. Another example that comes to mind is how women fighters were especially showcased by PYD/YPG and the Gulf countries as a source of propaganda and a counterbalance for legitimacy against well-publicized ISIS' negative attitude towards women (Szekeley, 2020).

Another topic of discussion is the role of the media when it comes to armed groups. Most of the recent research has focused on how certain armed groups are using social media to recruit their members, such as ISIS (Awan, 2017). Social media helps these groups to reduce the costs of dissemination and increase the information flow (Zeitsoff, 2017). While the role of social media is considered widely, the discussions on traditional news media remain limited, especially when social media shows that media itself is very effective in forming opinions about a certain armed group. Even when it comes to statecraft, the media becomes the way in which some ideas are conducted in society regarding foreign policy formation (Baum & Porter, 2008). While newspaper data is used for discerning motivations for the disarmament of armed groups (Malone, 2022), armed group-specific examination of newspapers (Aboud et al., 2023; Perkoski, 2019), there is not much interest in how armed groups themselves are depicted in the news media. While there are some examples of the framing of terror attacks, such as the attack committed by Anders Breivik (Falkheimer& Olsson, 2014) and the framing of the War on Terror and Islam in relation to terrorism (Woods, 2011), there are not necessarily any studies that delve into effects the ideological differences might have on these framings. Similarly, women in the media, when it comes to these armed groups, are widely discussed; however, these studies also focus on more specific cases, such as women joining ISIS (Martini, 2018) or the main narratives of women in media in relation to the groups they join (Nacos, 2005); not necessarily using the ideological background as a discussion point on media depiction. Exploring some general trends these groups have in relation to their ideology can help to further elaborate on how women are depicted in the news media through computational text analysis methods in a comparative manner. Therefore, such research can pave the way to examining multiple armed groups’ depictions based on their ideological difference. To fill this gap, the paper aims to explore how different ideological backgrounds can affect the media depiction of women in such contexts.

**Data Collection**

For this inquiry based on Lutz & Lutz (2019), two prominent types of armed groups were chosen: ethnic and religious. Four different groups were selected from these categories to offer a diverse range of candidates. Among religious groups, two very popular groups within the news media, Hamas and Hezbollah, were chosen. While both of the groups are Muslim and have fought against Israel since their foundation, they are distinguished by their sect, with Hamas being a Sunni and Hezbollah being a Shia organization. Among ethnic armed groups, two prominent groups in the news media from different continents were chosen: the Liberation Tigers of Tamil Eelam from Sri Lanka and the Provisional Irish Republican Army from Northern Ireland.

Table 1 - Group Typologies and search queries

| Name of the Group | Group Typology | Search Query | Search Years |
| --- | --- | --- | --- |
| Hamas | Religious | “Hamas” | 01.01.1987  31.12.2010 |
| Hezbollah | Religious | "Hizbullah" OR "Hezbollah” | 01.01.1984  31.12.2010 |
| Liberation Tigers of Tamil Eelam (LTTE) | Ethnic | “Tamil Tigers OR LTTE” | 01.01.1984  31.12.2009 |
| Provisional Irish Republican Army (IRA) | Ethnic | “Irish Republican Army” | 01.01.1980  31.12.2005 |

The data was taken from the newspaper news released when the groups chosen for the data categories were active. NexisUni was utilized to collect this news, which provides access to historical newspapers with 100 downloads of full text in a batch and 2500 downloads per day. To download these pieces of news, three parameters were sought: 1) The date of the search for the news was limited to before 2010, considering that the ethnic armed groups of concern were disbanded by 2010 (LTTE in 2009 and Provisional IRA in 2005) and nature of the news might have changed during the post-2010 period. 2) The parameters in Table1 were used specifically, including the name of the organization and the search query "women or female" to ensure that the news was mentioning about women in one way or another, the earliest mention of the said organization until 2010 or the year the organization ceased its activities, excluding the mentions of "men and women" since it is often used as another method of saying "people". 3) Within religious groups, since there was too much news in the other types of media, these were limited to newspapers. After data cleaning and removing duplicate news, the overall number of news was 6459 for Hamas, 2840 for Hezbollah, 2872 for LTTE, and 2572 for IRA, with 14743 news in total. While the source of the news was considered to be very valuable for an additional comparison (i.e., for example, Associated Press, Times of Israel, etc.), especially since they might skew the discourse, they were ultimately not included not to complicate the research design. However, the sources of the news were considered with most of the news being originated from the Western sources. The exceptions were the Jerusalem Post (284 news out of 9299 news) and United Press International (UPI)[[1]](#footnote-2) for ethnic news; only LTTE had a wide coverage from non-Western sources with the local media (Wijeya Newspaper Ltd. and the Associated Newspapers of Ceylon Ltd.)and the media from India (HT Media Ltd.). However, removing these news sources did not change the sentiment relations in all three cases, therefore the results were considered with all news sources combined. Therefore, as discussed later, these results should be considered for media that is 1) published in English 2) with a Western origin. While headlines were included in the data preprocessing, they were not used since some of them were absent. However, such inquiry might be considered in the future.

Table 2 - Most frequent news sources for religious and ethnic armed groups

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Religious | Country | News count | Ethnic | Country | News count |
| Nationwide News Pty Limited | Australia | 801 | Associated Press | United States | 1261 |
| CanWest Interactive | Canada | 441 | Nationwide News Pty Limited | Australia | 368 |
| The New York Times Company | United States | 418 | Agence France Presse | France | 312 |
| Guardian Newspapers Limited | United Kingdom | 308 | U.P.I. | Various | 307 |
| The Jerusalem Post | Israel | 284 | Deutsche Presse-Agentur | Germany | 216 |
| Times Newspapers Limited | United Kingdom | 279 | Wijeya Newspaper Ltd. | Sri Lanka | 201 |
| International Herald Tribune | France | 258 | HT Media Ltd. | India | 189 |
| Metroland Media Group Ltd | Canada | 255 | The New York Times Company | United States | 188 |
| Independent Print Ltd | United Kingdom | 251 | Press News Limited, | New Zealand | 174 |
| St. Louis Post-Dispatch, Inc. | United States | 245 | The Associated Newspapers of Ceylon Ltd. | Sri Lanka | 171 |

**Methodology**

This paper examined the research question in three steps. First, the news collected was explored and examined to infer their content in general. This step was particularly important since it aided the research in knowing what the dataset contained for the analysis purposes and to have an idea of what kind of news women were mentioned in. After getting an idea of the news, the tone of the news, based on topics, was examined since the tonality of the news affects both the impression the audience gets from reading such news and the depiction of the armed groups in relation to women. While these steps told a large portion of what the news that women were in, an additional step was employed to see exactly how women were depicted in these pieces of news and what their position in the news media was. To do so, dependency parsing was employed, which separates the words based on their grammatical relationship.

Different approaches were taken into consideration to preprocess the data. Except for removing unnecessary parts within a text (line breaks, page numbers, headers, and footnotes) and removing text that was larger than 1000 words (since long texts might skew the topic modeling and they are likely to contain irrelevant text such as news transcripts that contain various subjects other than armed groups), no particular method of preprocessing was employed for sentiment analysis and dependency parsing since the models used for such tasks included their own preprocessing methods and as the optimal methods in which these were trained should be kept since it is difficult to evaluate the performance of the methods on this dataset without any coded data.[[2]](#footnote-3)

For the first task, namely exploring and inferring the content of the news, topic modeling methods, i.e., statistical models that discern and separate topics based on hidden patterns, were decided. Based on the distinction made by Vayansky & Kumar (2020), ultimately, LDA was chosen as it was expected that the newspaper articles would be longer than 50 words, which eliminated the short text alternatives, and the time dimension was not found necessary. Among the remaining options CTM (Correlated Topic Model)[[3]](#footnote-4) (Blei & Lafferty, 2007) and LDA (Latent Dirichlet Allocation) (Blei et al, 2003), CTM is more resource and time-intensive, and correlation between the topics was not seen as important for the task at hand due to the distinct topics sought by the research goals. For this reason, it was also not considered. Hence, LDA was seen as sufficient enough as the main topic modeling method. The gensim library (Řehůřek & Sojka, 2011) was utilized to conduct topic modeling. For preprocessing, the data was cleaned within this order: lowercasing, lemmatization, removal of punctuation and stopwords, the armed group names and search queries, ethnic groups and places associated with the conflict, and some repeated words[[4]](#footnote-5). In addition, the armed group names and search queries were removed for topic modeling to prevent clustering over these common words. The data was divided into two based on the ideological typology, namely ethnic and religious armed groups, and were trained afterward). The model was trained with 50 iterations and 5-10 topics.[[5]](#footnote-6) Afterwards, the models were evaluated based on their coherence and perplexity, and finally, the topics were looked through with a human eye to see if they were interpretable. UMass[[6]](#footnote-7) (Minmo et al., 2011) and perplexity were used to measure coherence and perplexity since they were designed specifically to evaluate LDA topics. In addition, Cv coherence (Röder et al., 2015), a very popular measure for LDA, was also used for evaluation of the model to account for biases that UMass measure might have. Lower perplexity, higher coherence score for Cv and lower coherence score indicated a good fit for the LDA model. After checking the best performing topic statistically, the topics were also interpreted by the author by looking at the 30 articles that contains the highest score for each topic for each topic.

After labeling the text based on the topics, sentiment analysis was conducted on the news separated based on the LDA topics (to do so, only the topic values larger than 50% were used) to see how the media portrayal differed between different topics regarding armed groups and women. Over here, a choice between a supervised approach and a lexicon-based approach was considered. While supervised methods are more accurate, with transformer-based models such as BERT showing accuracy above 90% (for example, movie reviews (Alaparthi& Mishra, 2021)), training data for news data are mostly limited to news related to economics and finance (Sousa et al., 2019), news classification tasks (Del Corso et al., 2005) or are based on shorter texts such as news headlines (Malo et al, 2014). In addition, applying these methods might be computationally intensive. On the other hand, while lexicon-based analyses are lower in accuracy, they are easier to apply since they require no training data, are computationally less costly, and the data at hand use standardized English, unlike social media data. In addition, since as long as data is large, presuming that the errors are random, accuracy, while important, should not pose a significant issue for the task. Considering these options, three models were chosen. Two transformer-based pre-trained models, SieBERT, (Hartmann et al., 2022) and DistilRoberta-financial-sentiment (Romero, 2024)) were chosen as SieBERT, a RoBERTa model trained for sentiment analysis, and was trained on different types of the dataset including reviews, political tweets, etc. as a comprehensive model, while DistilRoberta-financial-sentiment was chosen it is trained on financial news to see if the news as a text format matters in capturing the sentiment in a similar context. In addition, since it is a lighter version of RoBERTa, a lighter version working well, would help the researchers with less computational capacities.[[7]](#footnote-8) For a lexicon-based approach, VADER (Hutto & Gilbert,2014) was chosen as the main model since it outperforms other popular lexicon-based methods (in some cases, even traditional ML methods) in editorials, which are close to news media[[8]](#footnote-9). One concern here is the inherent negative nature of the news about armed groups. There is a comparison considered for this task, which is randomly selecting news from the entire sample and comparing it with the sample score (quasi-randomization) and testing the difference with a t-test, which can also help to alleviate the potential lack of accuracy of the methods at hand.[[9]](#footnote-10) 5% of the data (737 random news texts) was randomly chosen to be the control group and trained with the respective model that it was tested with. Due to the lack of human coded data on news data about war, as far as the author’s knowledge goes, therefore, the metrics such as F1 scores were not used. Both SieBERT and DistilRoBERTa-financial-sentiment provide two values: one of the types of sentiments and the second the polarity of the said sentiment (e.g., Positive, 0.853). To get a continuous measure from these values, the neutral values were taken as 0 regardless of the polarity score[[10]](#footnote-11), and the positive polarities were kept as such and the negative ones were negated. Unlike these, VADER provides a composite continuous score that is negative (-1 to -0.5), neutral (-0.5 to 0.5), positive (0.5 to 1). In addition, for SieBERT, the texts were truncated to 512 tokens because of the requirements of the model. The preprocessing steps were skipped for these data since VADER uses its own preprocessing method to calculate sentiments that include punctuations, uppercasing, and as such, while the default preprocessing methods were opted for in the case of SieBERT and DistilRoBERTa.

One goal of the research is to discern which actors of importance are associated with which contexts (in this case, women) and actions in order to draw more details. Therefore, dependency parsing was used to extract the context in which these actors actively and passively participated in actions under which contexts that were explicitly mentioned by extracting sentences. Hence the sentences that contained the words (women, female, she, her, hers, herself) were chosen for this task. For this particular purpose, three packages were considered: SpaCy (Montani et al., 2023), Stanza (Qi et al, 2020), and NLTK (Loper& Bird, 2002). While SpaCy provides a more user-friendly environment than the others, the developers of the library do not recommend the use of SpaCy for research purposes, while NLTK remains a popular and flexible option. StanfordCoreNLP's dependency parsing model in Stanza was chosen for this task as it is trained based on the Penn Treebank (Marcus et al., 1993), which is based on newspaper text and is a widely used dataset for dependency parsing. After parsing, the top 50 most frequent words (with the removal of redundant words) associated with women were shared. A topic distinction was not made since these were intended to check how the media treated women in relation to different types of armed groups. The model was applied without any preprocessing (except removal of unnecessary ones such as line breaks, etc.) as the model used to calculate the dependencies by using the punctuation, stopwords, and uppercase. After the analysis, these were removed except "she, her, hers, and herself" and were lemmatized for a better comparison. In this analysis, both sides of the dependency of the word "woman" were employed. However, for "female," only the cases in which it was only used as a dependency since female" as an adjective could explain the roles of women the best instead of female being the root of the dependency.

**Results**

Six topics were chosen in both cases since they were interpretable by the human eye, and had the closest values to 0 in both coherence and perplexity (Appendix C) for both ethnic and religious groups.[[11]](#footnote-12) Within these topics, only one topic was fully about the LTTE. The rest were for religious groups, the news about Islam, and the Israeli-Palestinian conflict, conflict in general, international reactions to the conflict, and armed group's actions as a form of terrorism or as a crime. Within the ethnic groups' topics, there were topics on the international reaction against the LTTE, the attacks against civilians in general, bombings (often suicide), the peace process of IRA and LTTE, and the conflict between the combatants (Table 2 and 3).

Table 3 - Output of LDA for religious groups (most frequent 10 words were shown here)

| Religious news | War against Israel | Conflict news and condition of civilians | Irrelevant news | International reaction | Terrorism as crime |
| --- | --- | --- | --- | --- | --- |
| one, many, war, even, like, world, child, muslim, time, al | killed, attack, yesterday, militant, city, arafat, one, group, official, strip | rocket, military, attack, one, killed, war, civilian, soldier, force, fire | new, one, yesterday, government,date, load, day, official,killed,three | peace, state, president, minister, government, leader, bush, one, date, new | yesterday, killed, official, border, egyptian, egypt, attack, one,date, minister |

Table 4 - Output of LDA for ethnic groups (most frequent 10 words were shown here)

| International response against LTTE | Attacks against civilians | Bombings | Peace process | Military conflict | Irrelevant news |
| --- | --- | --- | --- | --- | --- |
| government ,civilian, rebel, killed, military, force, war, attack, international, area | government, peace, group, one, bomb, killed ,political ,last, party, three | bomb, attack, government ,omagh, minister, group, killed, load, security, one | peace, government, party, minister, one, date, leader, load, today, new | government, rebel, attack, killed, military, civilian, one, area, war, bomb | new, party, one, first, president, government ,day ,date ,day, last |

Both SieBERT and VADER analysis tools showed similar results except VADER having a more negative inclination. SieBERT did not label any neutral values while DistilBerta coded almost 75% of the data in religious news and 73% of the ethnic news data as neutral. When applied to each topic, most results were the same except for the peace process news and international reaction towards LTTE, with VADER showing a negative relation and SieBERT showing a positive one. News related to the conflict and civilian deaths were more negative than the control group in both cases, while international reaction and peace process were taken as more positive. In comparison to the ethnic groups, religious groups had more negative news, with SieBERT only showing significant results. Generally, the news about conflict was mostly negative, while international intervention and civilian life was taken in a positive light.

Table 5 -Summary statistics of different sentiment analysis methods

| Measures | VADER | SieBERT | DistilRoBERTa |
| --- | --- | --- | --- |
| Median | -0.9866 | -0.9808 | 0 |
| Mean | -0.5889 | -0.1249 | -0.1531 |

Table 6- Comparison of religious and ethnic news sentiments

| Model | t-statistics | p-value |
| --- | --- | --- |
| VADER | -1.027 | 0.3041 |
| SieBERT | -14.2477 | 0.00\*\*\* |
| DistilRoberta-financial-sentiment | -0.3326 | 0.7393 |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

When it comes to the sentences related to women, women are usually depicted with words such as different forms of violent actions e.g various conjugations of the verb "to kill,"" to wound," and "to die." In addition, some of the words referred to vulnerability, such as different conjugations of "old," "innocent," "pregnant," and "prison." The most commonly associated words with all of the cases were "men", "women," and "children," with men most likely referring to "men and women" and children being a coupling of "women and children" as vulnerable groups. In both cases, the identity of the women was prominent, such as "Palestinian" in Hezbollah and Hamas, "Lebanese" and "Christian" in Hezbollah, "Muslim," "Israeli" and "Jewish" in both, and foreign women such as "American, British, French, Iranian." Depictions of the attire of women (wear, dress)[[12]](#footnote-13) were also prevalent. The dresses were primarily used to depict identity, especially prevalent among the news related to religious armed groups with examples of Muslim attire such as the veil (Appendix B).

Dependencies of the word "female" were more divided, with suicide bombers being very relevant for religious groups and LTTE and many civilian professions being mentioned in the case of IRA and religious groups. While every group mentioned "female fighters" or similar combinations, only religious groups had suicide bombers, prisoners, and civilian and military personnel altogether. This suggests that the news about religious groups has a more composite and polarized nature, featuring women at each side of the spectrum both as a victim and as an active participant.

As the results suggest, women are depicted as victims, which is in line with them being a vulnerable group, often subject to killing, dislocation, and other forms of war crimes such as sexual assault. However, one interesting aspect is the depiction of women, often with children, adding to the narrative surrounding women as a vulnerable group. Another prominent factor is that, except for the IRA, they appear as a popular choice for suicide bombing as well. Ethnic groups do not necessarily show a distinction as in LTTE news; bomber ranks the highest (3rd most common word in relation to female). In other words, women from Hamas and Hezbollah are depicted as suicide bombers as well.

**Discussion**

While killings of women and conflict were very much prominent in both types of news, the news about religious armed groups contained more frequent words about killing than ethnic groups.[[13]](#footnote-14) Most of these results were also in line with sentiment analysis results conducted here (see Appendix A, Table 5), showing religious group news being more negative than ethnic group news content-wise. The news that mentioned women also included news of violence, war and attack on civilians. Interestingly, the news also involved a much internationalized dimension, confirming internationally the relation between armed groups and women were discussed not necessarily as a domestic condition but an internationalized phenomenon.

Depiction of women in both of these groups commonly include women as a vulnerable group often coupled with other vulnerable groups, either as victims or prisoners. In addition, they appear as suicide bombers in most cases except the IRA. In addition, women as combatants or supporters of the groups appear as a common frame. References to identity are very common, oftentimes nationality and religion is invoked in both cases, regardless of the role of the group. However, in this case, the groups differ in the sense that while ethnic group news is divided when it comes to which role women have, religious group news mentions women in all of these roles and with very specific references to the clothing unlike more generic references to the clothing in the ethnic groups.

Methodology-wise this research shows that when it comes to the conflict/political news, the content of the data matters (built on financial news) more than the structure of the data (as news), as both VADER and SieBERT models gave similar results. Except for a few cases, this suggests that VADER is more sensitive towards negativity than SieBERT. While it is difficult to suggest if one performed better than the other, there are some considerations that can be drawn from this comparison. Presuming that news about peace related to LTTE should be positive, VADER might have failed to capture this unlike SieBERT. While it is considered that the length of the text might have affected the difference in the result given, when the results are reconsidered with 400 words, which approximates the 512 token limit of BERT, the results are the same for these cases (except higher *p* values because of the smaller sample size[[14]](#footnote-15)). One potential explanation is that while VADER measures the tonality of the words within the text, SieBERT might give results regarding the meaning of the text because it is trained on human-coded data.[[15]](#footnote-16) In addition, VADER might have a slight negativity bias. Therefore, for such an endeavor, both have their use, with VADER being more useful for discerning the tonality of the language used while SieBERT discerning the nature of the content itself. General models like SieBERT and VADER performed well enough for both tasks since, for some obviously negative results, such as news about war, they managed to show negative sentiments.

In general, this research provides a preliminary framework for analyzing news in relation to armed groups. It confirms that news media also shows the vulnerability of women in conflict scenarios and their participation in the conflict in different manners. Most depictions of women are in conflict scenarios, while women are also mentioned in international reactions and civilian life during the conflict. In addition, some further research agendas were opened with this research, such as why religious groups hold a more negative, mixed, and polarized depiction of women (multiple depictions of women with clashing narratives) than the ethnic groups. In addition, this news can be compared with the groups' manifestos.

**Limitations**

However, there were some limitations to the study. First of all, none of the data was human-coded for sentiment due to the time limitations of the research. This made the use of popular metrics such as F1 not possible. To the knowledge of the author, a dataset that specifically has human-coded sentiments for political/conflict news does not exist. Although two of the methods (VADER and SieBERT) were tested to be effective enough for sentiment analysis tasks by the authors, this remains a deficit of the research. Therefore, the evaluation of the methods used remained limited to comparison. The data itself mostly comes from Western sources, so these results should be considered as the perspective of popular Western news sources such as the Associated Press, Guardian, etc., that work with NexisUni. To generalize this further, additional data from other news source databases can be collected. From another perspective, since women were often mentioned in a wider context within the news and specifically mentioned regarding the main events in a limited dataset, future analysis should include more distinctions, such as left/right-wing organizations and more organizations in general, to provide a better comparison for news framing. Similarly, since the nature of journalism has moved towards a more online direction with the advent of the Internet, the results found here might differ in today's world. Finally, while this study provides an exploration of the similarities and differences in the depiction of women in news about armed groups of different natures, a more qualitative analysis should be considered in the future, asking why such differences exist.

**Conclusion**

Overall results suggest that religious groups are depicted in a mixed manner and often more negative messages qualitatively than ethnic groups. While ethnic groups also have some properties of religious groups, the religious groups are depicted in a very polarized manner (subordinated women versus women as fighters) with more attention to their clothing, with suicide bombing being a very common role given to women. More generally, the results suggest that women are often seen as a vulnerable group that is killed, imprisoned, etc.. The results also suggest that there is reportage of international attention to these issues. While groups find it important, at least by 2010, it can be easily suggested that the media depict the condition of women in the news about the armed groups in a very negative light, often under news with highly negative sentiments. This paper also opens up further computational analysis of armed groups in the media with a more extensive comparison.

**References**:

Aboud, N., Berlin, M., Biasi, S., & Parker, T. B. (2023). Frame your friends close and your enemies closer: how al-Masra newspaper imagines individuals. *Behavioral Sciences of Terrorism and Political Aggression*, 1-21.

Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. *Journal of Marketing Analytics*, *9*(2), 118-126.

Awan, I. (2017). Cyber-extremism: Isis and the power of social media. *Society*, *54*(2), 138-149.

Bangerter, O. (2011). Reasons why armed groups choose to respect international humanitarian law or not. *Int’l Rev. Red Cross*, 93, 353.

Baum, M. A., & Potter, P. B. (2008). The relationships between mass media, public opinion, and foreign policy: Toward a theoretical synthesis. *Annu. Rev. Polit. Sci.*, *11*, 39-65.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., &Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, *5*(4), 323-370.

Blei, D. M.,&Lafferty, J. D. (2007). A correlated topic model of science.*Ann. Appl. Stat. 1 (1)* 17 - 35,.<https://doi.org/10.1214/07-AOAS114>

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of verbal learning and verbal behavior*, *8*(1), 1-8.

Cockburn, C. (2013). War and security, women and gender: an overview of the issues.

*Gender & Development*, 21(3), 433–452.

Cohen, D. K. (2017). The ties that bind: How armed groups use violence to socialize fighters. *Journal of Peace Research*, *54*(5), 701–714.

Del Corso, G. M., Gulli, A., & Romani, F. (2005). Ranking a stream of news. In *Proceedings of the 14th International Conference on World Wide Web* (pp. 97-106).

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ...& Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the national academy of sciences*, *112*(8), 2389-2394.

Falkheimer, J., & Olsson, E. K. (2015). Depoliticizing terror: The news framing of the terrorist attacks in Norway, 22 July 2011. *Media, War &Conflict*, *8*(1), 70-85.

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*, *40*(1), 75-87.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Loper, E., & Bird, S. (2002). Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Lutz, J., & Lutz, B. (2019). *Global terrorism*. Routledge.

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., &Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, *65*(4), 782-796.

Malone, I. (2022). Unmasking militants: Organizational trends in armed groups, 1970–2012. *International Studies Quarterly*, *66*(3)

Marcus, M., Santorini, B., &Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics*, *19*(2), 313-330.

Manekin, D., & Wood, R. M. (2020). Framing the Narrative: Female Fighters, External Audience Attitudes, and Transnational Support for Armed Rebellions. *Journal of Conflict Resolution*, *64*(9), 1638-1665. https://doi.org/10.1177/0022002720912823

Martini, A. (2018). Making women terrorists into “Jihadi brides”: An analysis of media narratives on women joining ISIS. *Critical Studies on Terrorism*, *11*(3), 458-477.

Mimno, D.,Wallach, H., Talley, E., Leenders, M., &McCallum, A. (2011, July). Optimizing semanticcoherence in topicmodels. In *Proceedings of the 2011 conference on empirical methods in natural language processing*(pp. 262-272).

Montani, I., Honnibal, M., Honnibal, M., Boyd, A., Van Landeghem, S., & Peters, H. (2023). *explosion/spaCy: v3.7.2: Fixes for APIs and requirements*. doi:10.5281/ZENODO.1212303

Nacos, B. L. (2005). The portrayal of female terrorists in the media: Similar framing patterns in the news coverage of women in politics and in terrorism. *Studies in conflict & terrorism*, *28*(5), 435-451.

Perkoski, E. (2019). Internal politics and the fragmentation of armed groups. *International Studies Quarterly*, *63*(4), 876-889.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Řehůřek, R., &Sojka, P. (2011). Gensim—statistical semantics in Python. *Retrieved from gensim. org*.

Romero, M. (2024). *DistilRoBERTa-financial-sentiment* [Software]. Hugging Face. <https://huggingface.co/mrm8488/DistilRoBERTa-finetuned-financial-news-sentiment-analysis>

Röder, M., Both, A., &Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).

San-Akca, B. (2016). *Statesin disguise: causes of state support for rebel groups*. Oxford

University Press.

Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019, November). BERT for stock market sentiment analysis. In *2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI)* (pp. 1597-1601). IEEE.

Szekely, O. (2020). Fighting about women: ideologies of gender in the Syrian civil war. *Journal of Global Security Studies*, *5*(3), 408–426.

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, 101582. URL: https://doi.org/10.1016/j.is.2020.101582

Woods, J. (2011). Framing terror: An experimental framing effects study of the perceived threat of terrorism. *Critical Studies on Terrorism*, *4*(2), 199-217.

**Appendix A – t-tests based on typology and topics**

Table 1 – VADER t-test results (Hamas)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 2.4340 | -11.5308 | -12.49 | -1.6412 | 12.1864 | -2.9925 |
| p-value | 0.019\*\* | 0.00\*\* | 0.00\*\*\* | 0.101 | 0.00\*\*\* | 0.003\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 2 – SieBERT t-test results (Hamas)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 1.8430 | -10.2310 | -7.8557 | 1.7980 | 10.5026 | -4.0615 |
| p-value | 0.065\* | 0.00\*\*\* | 0.00\*\*\* | 0.8595 | 0.00\*\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 3 – DistilRoBERTa-financial-sentiment t-test results (Hamas)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 8.6058 | -11.0181 | -6.3807 | -0.7125 | 10.2615 | -0.8509 |
| p-value | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.4763 | 0.00\*\*\* | 0.3951 |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 4 – VADER t-test results (Hezbollah)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 3.9678 | -10.7326 | -12.6134 | 0.1770 | 10.5026 | -4.0615 |
| p-value | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.8595 | 0.00\*\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 5 – SieBERT t-test results (Hezbollah)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 5.3731 | -5.3069 | -4.5176 | 3.0374 | 4.4211 | -0.1430 |
| p-value | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.0024\*\* | 0.00\*\*\* | 0.8863 |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 6 – DistilRoBERTa-financial-sentiment t-test results (Hezbollah)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 7.6417 | -6.9190 | -4.0012 | -0.7389 | 8.6150 | -1.5836 |
| p-value | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.4603 | 0.00\*\*\* | 0.11 |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 7 – VADER t-test results (LTTE)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | -1.9950 | 1.7672 | 6.4167 | -0.8184 | -10.9056 | 3.6231 |
| p-value | 0.0462\*\* | 0.0792\* | 0.00\*\*\* | 0.4141 | 0.00\*\*\* | 0.0003\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 8 –SieBERTt-test results (LTTE)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | 3.034 | 2.6246 | 6.5157 | 4.1021 | -2.071 | 10.5791 |
| p-value | 0.002\*\*\* | 0.01\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.039\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 9 – DistilRoBERTa-financial-sentiment t-test results (LTTE)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | -3.8651 | 4.9030 | 4.7970 | 2.89 | -10.8863 | 4.5308 |
| p-value | 0.0001\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.0042\*\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 10– VADER t-test results (IRA)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | No topic | 0.1201 | -12.7779 | 4.4435 | -10.095 | 9.2261 |
| p-value | No topic | 0.9045 | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 11 – SieBERT t-test results (IRA)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | No topic | 4.3274 | -2.2508 | 9.2688 | -4.91 | 10.0135 |
| p-value | No topic | 0.00\*\*\* | 0.025\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

Table 12 – DistilRoBERTa-financial-sentiment t-test results (IRA)

| Topics | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| --- | --- | --- | --- | --- | --- | --- |
| t-statistic | No topic | 3.9732 | -3.068 | 6.9572 | -4.5263 | 10.0748 |
| p-value | No topic | 0.00\*\*\* | 0.0024\*\*\* | 0.00\*\*\* | 0.00\*\*\* | 0.00\*\*\* |

p<0.1\*,p<0.05\*\*,p<0.01\*\*\*

**Appendix B - Results of Dependency Parsing**

Table 1 – Women as head of the dependency

| Hamas | Frequency | Hezbollah | Frequency | LTTE | Frequency | IRA | Frequency |
| --- | --- | --- | --- | --- | --- | --- | --- |
| men | 413 | men | 198 | men | 312 | men | 284 |
| killed | 361 | killed | 129 | child | 128 | woman | 114 |
| child | 263 | child | 105 | bomber | 95 | root | 108 |
| woman | 182 | woman | 91 | woman | 88 | people | 69 |
| root | 151 | root | 71 | civilian | 71 | said | 62 |
| killing | 123 | people | 65 | people | 55 | became | 35 |
| civilian | 119 | killing | 51 | root | 53 | group | 29 |
| people | 113 | civilian | 41 | cadre | 53 | killed | 29 |
| prisoner | 91 | said | 31 | soldier | 47 | child | 26 |
| said | 86 | wounded | 30 | said | 44 | lost | 25 |
| release | 75 | right | 29 | killed | 41 | soldier | 24 |
| palestinian | 73 | group | 27 | fighter | 41 | prisoner | 22 |
| bomber | 70 | number | 19 | wing | 33 | died | 21 |
| died | 69 | injured | 19 | included | 26 | right | 20 |
| hundred | 68 | prisoner | 17 | use | 25 | treated | 19 |
| soldier | 67 | shot | 17 | used | 24 | girl | 19 |
| group | 65 | hundred | 16 | group | 21 | man | 18 |
| wounded | 57 | soldier | 16 | rebel | 21 | marched | 17 |
| number | 51 | death | 16 | carried | 20 | number | 15 |
| shot | 49 | release | 16 | tamil | 19 | admitted | 14 |
| man | 49 | dozen | 15 | violence | 18 | joined | 14 |
| included | 47 | women's | 15 | number | 17 | injured | 13 |
| right | 45 | wounding | 14 | blew | 17 | arrested | 13 |
| used | 40 | kill | 14 | leader | 17 | killing | 13 |
| saw | 37 | allowed | 14 | body | 17 | tourist | 13 |
| kill | 35 | student | 13 | boy | 17 | figure | 13 |
| carried | 33 | died | 13 | thousand | 16 | think | 12 |
| dead | 31 | use | 11 | using | 16 | hundred | 12 |
| released | 31 | dead | 11 | constable | 16 | take | 12 |
| dozen | 31 | carried | 11 | killing | 15 | member | 12 |
| screamed | 30 | role | 11 | arrested | 15 | suffered | 12 |
| left | 29 | wear | 11 | refugee | 14 | released | 11 |
| sat | 29 | left | 10 | victim | 14 | called | 11 |
| role | 28 | many | 10 | rape | 14 | scent | 11 |
| crowd | 28 | lebanese | 10 | recruiting | 14 | guerrilla | 10 |
| free | 27 | bomber | 10 | girl | 12 | told | 10 |
| israeli | 26 | attack | 10 | tiger | 12 | wounded | 10 |
| using | 25 | kissing | 10 | suicide | 12 | attack | 10 |
| arrested | 24 | palestinian | 9 | identified | 12 | say | 10 |
| wounding | 23 | get | 9 | right | 12 | gave | 10 |
| use | 23 | brought | 9 | raped | 11 | role | 10 |
| murder | 23 | wailed | 9 | dysentery | 11 | men's | 10 |
| body | 22 | victim | 9 | detonated | 11 | civilian | 9 |
| told | 22 | came | 9 | turn | 11 | thousand | 9 |
| injured | 22 | serviceman | 9 | ward | 11 | found | 9 |
| girl | 22 | likened | 9 | dead | 10 | casket | 9 |
| student | 21 | saw | 8 | person | 10 | included | 9 |
| cry | 21 | murder | 8 | life | 10 | officer | 9 |
| candidate | 20 | treat | 8 | sought | 10 | victim | 9 |
| include | 20 | included | 8 | died | 10 | coalition | 9 |

Table 2 - Women as the dependent of dependency

| Hamas | Frequency | Hezbollah | Frequency | LTTE | Frequency | IRA | Frequency |
| --- | --- | --- | --- | --- | --- | --- | --- |
| child | 1298 | child | 506 | child | 418 | child | 204 |
| palestinian\* | 615 | young | 100 | include | 223 | young | 116 |
| include | 314 | include | 98 | woman | 88 | woman | 114 |
| israeli\* | 243 | woman | 91 | young | 82 | man | 104 |
| woman | 182 | lebanese\* | 89 | tamil\* | 65 | protestant\* | 74 |
| young | 181 | palestinian\* | 76 | pregnant | 64 | pregnant | 72 |
| man | 154 | old | 73 | many | 61 | first | 68 |
| many | 143 | muslim\* | 66 | man | 56 | old | 60 |
| old | 140 | many | 57 | girl | 47 | include | 55 |
| muslim\* | 94 | israeli\* | 51 | elderly | 21 | mostly | 33 |
| elderly | 91 | man | 47 | old | 21 | many | 32 |
| innocent | 83 | mostly | 39 | lankan\* | 19 | black\* | 27 |
| pregnant | 74 | innocent | 38 | another | 17 | catholic\* | 27 |
| gaza | 61 | pregnant | 29 | especially | 15 | ireland\* | 26 |
| mostly | 47 | among | 25 | muslim\* | 15 | receive | 22 |
| arab\* | 47 | wear | 24+ | person | 15 | work | 19 |
| dead | 46 | elderly | 24 | mostly | 14 | together | 17 |
| among | 45 | work | 23 | baby | 13 | along | 16 |
| first | 45 | dress | 22+ | aged | 13 | arrest | 16 |
| wear | 41+ | mainly | 21 | rebel | 12 | belfast | 15 |
| kill | 41 | kill | 19 | aboriginal\* | 12 | girl | 15 |
| hold | 40 | girl | 19 | attack | 12 | elderly | 14 |
| several | 37 | first | 19 | boy | 11 | boy | 14 |
| dress | 37+ | clad | 18+ | saris+ | 11+ | among | 14 |
| another | 34 | veiled | 18+ | palestinian\* | 11 | irish\* | 13 |
| jewish\* | 34 | arab | 16 | live | 10 | elect | 13 |
| carry | 33 | aged | 16 | first | 10 | dress+ | 13 |
| prisoner | 33 | christian\* | 16 | local | 10 | american\* | 12 |
| minor | 33 | british\* | 16 | carry | 10 | say | 12 |
| veiled | 32+ | iranian\* | 15 | handicapped | 9 | prison | 12 |
| girl | 30 | jewish\* | 14 | die | 9 | sit | 11 |
| jail | 30 | american\* | 13 | detonate | 9 | another | 10 |
| aged | 29 | dead | 12 | expert | 9 | year | 10 |
| half | 26 | behind | 11 | want | 9 | hold | 10 |
| person | 24 | hold | 11 | country | 8 | serve | 10 |
| palestine | 23 | carry | 11 | lanka\* | 8 | marry | 10 |
| behind | 22 | live | 11 | dress+ | 8 | die | 9 |
| society | 21 | person | 11 | heroic | 8 | white\* | 9 |
| age | 21 | die | 10 | mother | 7 | carry | 9 |
| bomber | 21 | veil+ | 10 | give | 7 | local | 9 |
| blow | 20 | iraqi\* | 10 | identify\* | 7 | several | 9 |
| imprison | 20 | especially | 9 | work | 7 | meet | 9 |
| dozen | 19 | cover | 9 | chechen\* | 7 | british\* | 9 |
| say | 19 | every | 9 | say | 7 | force | 9 |
| cent | 18 | third | 9 | asian\* | 7 | look | 9 |
| work | 18 | robe+ | 9 | wound | 7 | jail | 9 |
| take | 18 | weep | 9 | sinhalese\* | 7 | make | 9 |
| french\* | 18 | another | 9 | several | 7 | kill | 9 |
| identify\* | 18 | party | 9 | widow | 7 | lose | 9 |
| american\* | 18 | country | 8 | village | 7 | condition | 9 |

(Asterisk shows the words related to identity and plus is about the clothing of women)

Table 3 – Female as the head of dependency

| Hamas | Frequency | Hezbollah | Frequency | LTTE | Frequency | IRA | Frequency |
| --- | --- | --- | --- | --- | --- | --- | --- |
| bomber | 209! | soldier | 73! | bomber | 353! | female | 62 |
| female | 118 | bomber | 50! | cadre | 100! | male | 18 |
| prisoner | 102 | female | 48 | male | 78 | college | 18\* |
| soldier | 78! | employee | 14\* | female | 46 | officer | 13! |
| member | 38! | prisoner | 13 | rebel | 32! | prisoner | 11 |
| male | 36 | male | 12 | suicide | 30! | minority | 10 |
| student | 34\* | student | 12 | soldier | 27! | member | 10! |
| suicide | 26! | victim | 8 | fighter | 25! | teacher | 9\* |
| relative | 21 | teacher | 8\* | tiger | 25 | first | 8 |
| officer | 19! | minister | 7\* | age | 18 | passenger | 8\* |
| prime | 17\* | member | 7! | prime | 18 | friend | 8 |
| lawyer | 16\* | prime | 6\* | body | 17 | politician | 7\* |
| companion | 15! | officer | 5! | leader | 17! | prime | 7\* |
| activist | 15\* | leader | 5! | officer | 16! | worker | 7\* |
| civilian | 12\* | home | 5 | child | 15 | employee | 6 |
| candidate | 11 | relative | 4 | civilian | 14\* | president | 6 |
| worker | 11\* | circumcision | 4 | member | 13! | star | 6\* |
| assailant | 10! | first | 4 | terrorist | 13! | companion | 6 |
| detainee | 10\* | recruit | 4! | attacker | 10! | cadet | 5! |
| attacker | 9! | colleague | 4\* | attack | 9! | bomber | 5! |
| terrorist | 9! | speaker | 4\* | suspect | 8 | student | 5\* |
| minister | 8\* | player | 4\* | doctor | 7\* | fetus | 5 |
| employee | 8\* | artist | 3\* | mujahedeen | 7! | activist | 5\* |
| journalist | 8\* | bodyguard | 3! | journalist | 6\* | head | 4 |
| passenger | 8\* | doctor | 3\* | student | 6\* | victim | 4 |
| root | 8 | assailant | 3! | victim | 6 | circumcision | 4 |
| dancer | 7\* | root | 3 | employee | 5\* | voter | 4 |
| settler | 7\* | civilian | 3\* | tamil | 5 | fighter | 3! |
| protester | 7 | politician | 3\* | operative | 5! | bishop | 3\* |
| circumcision | 7 | aide | 3\* | root | 4 | aid | 3 |
| teacher | 7\* | share | 3 | force | 4! | driver | 3\* |
| inmate | 7 | mourner | 2 | guerrilla | 4! | impersonator | 3 |
| volunteer | 6\* | win | 2 | bomber's | 4! | candidate | 3\* |
| palestinian | 6 | commentator | 2 | said | 4 | mostly | 3 |
| mutilation | 6 | elderly | 2 | recruit | 4! | marshal | 3! |
| co-host | 6 | inmate | 2 | insurgent | 4! | lieutenant | 3! |
| president | 5 | passenger | 2\* | commando | 4! | agent | 3 |
| fighter | 5! | attendant | 2\* | men | 4 | deputy | 3\* |
| bystander | 5 | face | 2 | ward | 4\* | attorney | 3\* |
| elderly | 5 | claimant | 2 | constable | 4\* | medical | 3\* |
| professor | 5\* | strict | 2 | prisoner | 3 | defendant | 3 |
| friend | 5 | librarian | 2\* | unit | 3 | inmate | 3 |
| lawmaker | 5\* | candidate | 2 | checking | 3 | governor | 3\* |
| judge | 4\* | president | 2\* | accomplice | 3 | allowed | 2 |
| general | 4 | chief | 2 | volunteer | 3 | commodity | 2 |
| chancellor | 4\* | servant | 2\* | relative | 3 | journalist | 2\* |
| body | 4 | agent | 2! | minister | 3\* | attendant | 2 |
| vote | 4 | became | 2 | refugee | 3\* | soldier | 2! |
| first | 4 | guard | 2 | detained | 3 | win | 2 |
| supporter | 4 | accountant | 2 | passer | 3 | percent | 2 |

(Asterisk marks civilian roles, while exclamation mark marks the military/active roles)

**Appendix C – Evaluation Figures for LDA**

Figure 1 – Evaluation Graphic for Religious News (UMass)

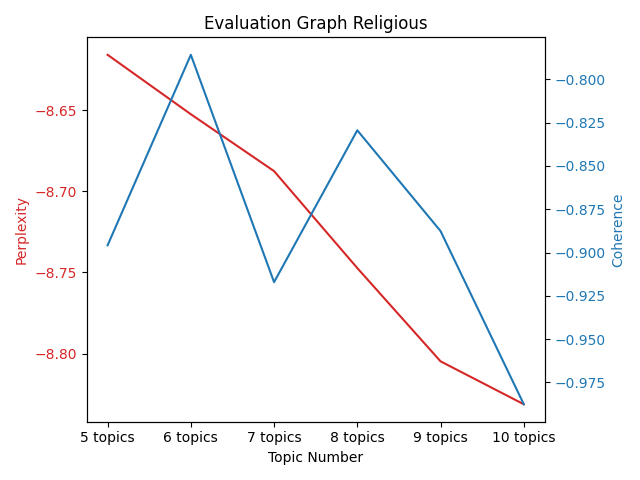


Figure 2 – Evaluation Graphic for Ethnic Armed Group News (UMass)

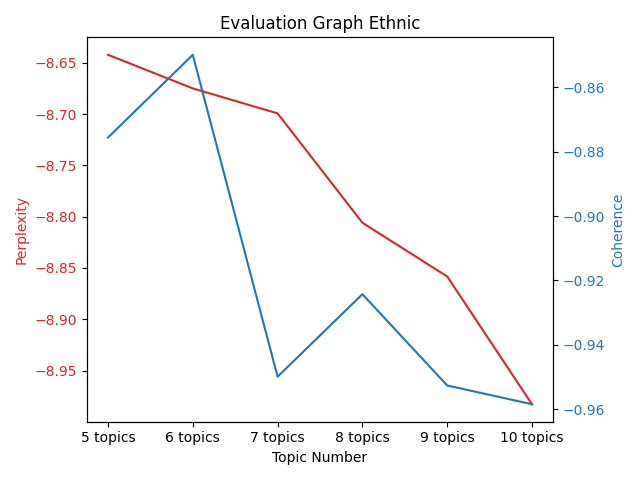


Figure 3 – Evaluation Graphic for Religious Armed Group News (Cv)

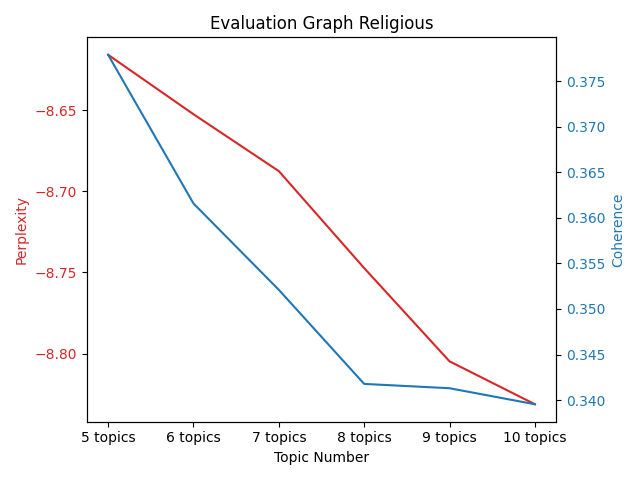
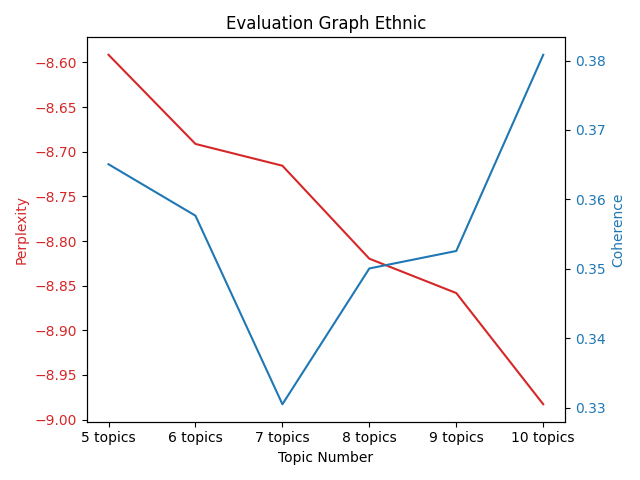


Figure 4 – Evaluation Graphic for Ethnic Armed Group News (Cv)



1. While UPI is a Western media company, the news sources they use include countries such as Iran, Pakistan, Qatar etc. [↑](#footnote-ref-2)
2. The reasons why preprocessing was not employed for these methods will be elaborated further in the discussion of the rest of the models. [↑](#footnote-ref-3)
3. The topic number is also expected to be smaller than 60. [↑](#footnote-ref-4)
4. These words were " would, two, ulster, year, de, la, sri, belfast, said, lanka, tamil, tiger, eelam, woman, female, sinnfein, irish republican army ,ap, ltte, tamil tigers, gaza, said, jerusalem, west bank, arab, palestinian, israel, hamas, us, army, police, lankan, people, mr, hezbollah, ira, ireland, irish, catholic, sinn, fein, protestant, republican, northern, israeli, british, northern ireland, united kingdom, lebanon, israel, palestine, wa, ha, also, west, bank" [↑](#footnote-ref-5)
5. This number was chosen due to the research's feasibility, the data's size, and ease of interpretation. [↑](#footnote-ref-6)
6. [↑](#footnote-ref-7)
7. DistilRoberta-financial-sentiment is expected to perform poorly since what counts as a negative in finance differs in comparison to what is negative under war contexts. However, it was still kept as an option to test if the model would perform well even if the content the model was trained on did not fit. Not to mention that not many sentiment analysis models are trained with political news data. [↑](#footnote-ref-8)
8. One caveat should be held in mind that VADER performed best with tweets and while it was getting the same results with the traditional methods of machine learning, the accuracy of these models were around 60% at best for labeling an editorial correctly. However, in large amounts of text, presuming the inaccuracies are randomly caused, this should not constitute an issue. [↑](#footnote-ref-9)
9. Two presumptions about the data were kept in mind. If the bias is random, the comparison between a control group and the actual dataset should produce valid results on the direction of the sentiment since the bias is random. If the biases are toward a certain direction, then this will not affect the relative direction of the text since both datasets have the same bias. In addition, the authors of the models SieBERT (Hartmann et al., 2022) and VADER (Hutto & Gilbert, 2014) have tested the measure on different forms of text, and the performance of the models deemed sufficient for a wide variety of use. [↑](#footnote-ref-10)
10. The reason for this choice is that determining the polarity of a neutral text can be challenging, as it is unclear whether the text leans towards a positive or negative sentiment if it was not a neutral text. [↑](#footnote-ref-11)
11. While the results suggest that 5 topics for religious and 10 topics for the ethnic groups are the optimal numbers, the second highest coherence is when there are 6 topics which also coincides with the results of UMass. While perplexity score gets lower with each topic, the sharpest decline happens when there are 8 topics. Considering that the coherence models point out to LDA performing poorly in topic number of 7, 6 was ultimately chosen. [↑](#footnote-ref-12)
12. There were cases in which traditional dresses such as sari were mentioned in the case of LTTE and the color green and berets in the case of the IRA. However, religious groups also include special verbs such as veiled, and the type of wear is a way more prominent in comparison to the ethnic ones. [↑](#footnote-ref-13)
13. Ranked 2nd and 6th/7th in religious groups and 11th and 8th in LTTE and IRA respectively. [↑](#footnote-ref-14)
14. The results for this analysis are not included in this paper but it can be easily replicated through using an already existing line in the code. [↑](#footnote-ref-15)
15. This might be an effect of the Pollyanna hypothesis, which proposes that people tend to use positive words and make positive meanings (Boucher & Osgood, 1969). This was also tested with textual data (Dodds et al., 2015). While this result might also be in line with aversion to bad impressions (Baumeister et al. 2021), such a conclusion needs more testing and is not in line with the purpose of this article. [↑](#footnote-ref-16)