

ARTICLE

The Gaza War: Text Analysis

Namkil Kang 

College of Liberal Arts, Far East University, Chungbuk, Republic of Korea

ABSTRACT

The objective of this paper is to conduct a text analysis of 40 BBC News broadcasts from the year 2024. Notably, the term “Gaza” appears with the highest frequency, totaling 335 tokens, and occupies a significant proportion of the coverage. This prominence is further illustrated by its substantial representation in a word cloud, suggesting its status as a pivotal topic. Additionally, the term “Israel” ranks second in prominence, indicating its relevance, though slightly less so than Gaza. Other noteworthy keywords include “homes,” “base,” “campaign,” “attack,” “conditions,” “group,” “leader,” and “ceasefire,” all of which contribute to the narrative surrounding the conflict. The data reveals that “Gaza” has a 10.5% likelihood of being identified as a main topic, while “Israel” has an 8.4% likelihood. Moreover, the similarity index indicates that the terms “Gaza” and “BBC” share the highest degree of similarity, followed closely by “Hamas” and “Gaza.” These findings underscore the importance of language in shaping public perception and discourse surrounding complex geopolitical issues. In conclusion, BBC News does not take a stance in favor of either Israel or Hamas; rather, it maintains a neutral position. This impartiality is effectively illustrated through various analytical methods, including term frequency analysis, word cloud analysis, network analysis, topic analysis, and similarity analysis.

Keywords: Big Data; Topic; Similarity; Term Frequency; Network; Word Cloud; Cluster

*CORRESPONDING AUTHOR:

Namkil Kang, College of Liberal Arts, Far East University, Chungbuk, Republic of Korea
Email: somerville@hanamail.net

ARTICLE INFO

Received: 29 October 2024 | Revised: 11 November 2024 | Accepted: 11 November 2024 | Published Online: 9 December 2024
DOI: <https://doi.org/10.30564/fls.v6i6.7638>

CITATION

Kang, N., 2024. The Gaza War: Text Analysis. *Forum for Linguistics Studies*. 6(6): 589-604. DOI: <https://doi.org/10.30564/fls.v6i6.7638>

COPYRIGHT

Copyright © 2024 by the author(s). Published by Bilingual Publishing Co. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

Gee ^[1] provides insights into discourse analysis and its ties to social practices. Biber, Conrad, and Reppen ^[2] focus on corpus linguistics to analyze language structure and usage. Halliday and Matthiessen ^[3] introduce systemic functional grammar, emphasizing context in text analysis. Swales and Feak ^[4] explore academic writing conventions and the analysis of scholarly texts. Van Dijk ^[5] addresses the cognitive and social dimensions of discourse, offering a comprehensive framework for linguistic analysis. Van Dijk ^[5] examines discourse as both structure and process, focusing on the cognitive and social dimensions of communication. Sampson ^[6] offers an empirical approach to linguistics, highlighting the importance of data-driven methods.

Our approach focuses on determining whether 40 pieces of BBC News align themselves with Israel or Hamas. This paper employs big data analysis techniques to conduct a text analysis of the Gaza war. It is important to acknowledge that war inherently results in profound tragedy. Utilizing Python for text analysis, this study aims to provide a comprehensive examination of these news pieces through six distinct analytical approaches.

First, we will explore the term frequency of relevant nouns identified within the articles. To ensure accuracy and relevance, adverbs, verbs, adjectives, prepositions, gerunds, to-infinitives, and relative pronouns have been excluded from this analysis. Focusing exclusively on nouns aligns with established methodologies in big data analysis. Term frequency refers to the number of specific nouns that appear across the analyzed documents, enabling us to identify the core vocabulary that shapes the narratives in these 40 pieces of news coverage.

Second, we will generate a word cloud that visually represents the core vocabulary identified within the articles. This visualization will help us capture and emphasize the most significant terms that constitute the thematic essence of the reports.

Third, we will analyze the networks formed by these relevant words, categorizing them into core and peripheral terms. Core terms will be positioned at the center of the network, while less significant terms will appear towards

the periphery. This network analysis will provide insights into the relationships between key concepts within the articles.

Fourth, we will examine clusters of interrelated words that demonstrate a high degree of correlation within the news coverage. Identifying these clusters will enable us to understand how certain terms are interconnected and contribute to the overall narrative of the Israel-Gaza conflict.

Fifth, we will assess the probability of specific terms becoming main topics within the analyzed news articles. These keywords will be evaluated based on their capacity to encapsulate the main themes of the documents, thereby informing us about which terms are most likely to serve as focal points in the discourse.

Finally, we will investigate the similarities between keywords to uncover the extent of their interrelatedness. Specifically, we will focus on identifying terms that are closely associated with “Hamas” and “Gaza,” respectively, thereby elucidating the nuanced connections between these critical concepts.

2. Materials and Methods

This study employs a systematic approach to analyze 40 BBC news articles pertaining to the Israel-Gaza war broadcasted in 2024. The methods employed in this analysis can be categorized into six key components, each aimed at extracting meaningful insights from the data.

2.1. Data Collection

The dataset consists of 40 articles from BBC news that were selected based on their relevance to the Israel-Gaza conflict during the year 2024. Each article was obtained from the BBC’s online archives, ensuring a comprehensive representation of the news coverage related to this ongoing issue.

2.2. Term Frequency Analysis

To identify the most frequently used nouns within 40 articles, we performed a term frequency analysis. Utilizing Python, we processed the text data to exclude non-

noun parts of speech, such as adverbs, verbs, adjectives, prepositions, gerunds, to-infinitives, and relative pronouns. The resulting dataset allowed us to focus solely on nouns, thereby enabling us to pinpoint the core vocabulary that characterizes the articles. The frequency of each noun was calculated and ranked to highlight the most significant terms.

2.3. Word Cloud Generation

Following the term frequency analysis, we generated a word cloud to visually represent the prominent nouns identified in the articles. This visualization was created using Python libraries such as Matplotlib and WordCloud. The size of each word in the cloud corresponds to its frequency, allowing for an immediate visual understanding of which terms are most central to the narrative.

2.4. Network Analysis

We conducted a network analysis to explore the relationships between the identified nouns. Using network analysis tools in Python, we classified the terms into core and peripheral categories based on their frequency and co-occurrence within the articles. Core terms were positioned centrally within the network diagram, while peripheral terms were displayed on the outskirts. This analysis facilitated a clearer understanding of how key concepts are interconnected within the discourse.

2.5. Cluster Analysis

To identify groups of interrelated words, we implemented a cluster analysis. This process involved calculating the similarity between nouns based on their co-occurrence in the articles. We utilized clustering algorithms available in Python libraries such as Scikit-learn to group nouns that share a high degree of correlation. This approach allowed us to discern thematic clusters that emerge from the news coverage.

2.6. Main Topic Probability Assessment

We assessed the probability of nouns becoming main topics by calculating the likelihood of each term being

categorized as a central theme within the articles. This was achieved by analyzing the frequency of each term in conjunction with its contextual relevance across the 40 documents. The terms with the highest probabilities were identified as potential main topics that encapsulate the core themes of the news coverage.

2.7. Similarity Analysis

Finally, we conducted a similarity analysis to evaluate the relationships between specific keywords, particularly focusing on “*Hamas*” and “*Gaza*.” By utilizing cosine similarity measures, we quantified the degree of similarity between various terms based on their co-occurrence patterns within the articles. This analysis provided insight into how closely these critical terms are related within the context of the coverage.

3. Results

3.1. Samples

In this section, we present a selection of 40 samples derived from BBC news articles related to the Israel-Gaza conflict. Each sample consists of the first sentence from one of the 40 documents analyzed. As previously outlined, we undertook a rigorous preprocessing step to enhance the clarity and relevance of our data. Specifically, we eliminated non-noun elements from the original texts, including adverbs, verbs, adjectives, prepositions, relative pronouns, to-infinitives, pronouns, and gerunds. This focused approach allowed us to concentrate exclusively on the nouns, which are critical for our subsequent analyses.

Table 1 displays the first sentence of each of the 40 pieces of BBC news, showcasing the context and language used in these reports. This selection provides a representative snapshot of the coverage and thematic elements presented in the articles.

Table 1. Samples.

Number	Sample
1	Walking through the ruins of what used to be h...
2	The first aid in two weeks has gone into north...
3	The Disasters Emergency Committee (DEC) is lau...
4	The US has written to Israel, giving it 30 days...

5	Witnesses to an Israeli air strike and resulti...
6	The first aid in two weeks has gone into north...
7	Britain took control of the area known as Pale...
8	Within days of Israel launching its war in Gaz...
9	Three months after a 13-year-old girl left Gaz...
10	A United Nations commission of inquiry has acc...
11	At least 28 people have been killed and 54 inj...
12	US President Joe Biden and Israel's Prime Mini...
13	A year ago, after the October 7 attacks and th...
14	Israel has held ceremonies to remember the vic...
15	Sir Keir Starmer marked "a day of sorrow and g...
16	According to Israel, 251 Israelis and foreigne...
17	It is one year since Israel responded to Hamas...
18	Millions of people in the Middle East dream of...
19	Following the strikes in central Gaza, videos v...
20	A few metres from a charred home in Kibbutz Be...
21	A retired surgeon has described dealing with d...
22	One year on from the 7 October Hamas attacks, ...
23	The most senior leader of Hamas outside Gaza h...
24	A Yazidi woman who was kidnapped aged 11 in Ir...
25	An Israeli air strike that killed multiple chi...
26	At the Sahara Restaurant in Dearborn, Michigan...
27	Up to 30 sick children are to be brought from ...
28	Tens of thousands of people have rallied acros...
29	The first full day of a campaign to vaccinate ...
30	A series of "humanitarian pauses" to fighting ...
31	A blogger who attracted a following on Instagr...
32	A medic from Greater Manchester has described ...
33	Kaid Farhan Elkadi, 52, is from a Bedouin vill...
34	The waters on parts of Gaza's Mediterranean co...
35	Earlier this week, on live television, the mot...
36	According to the UN, Gaza, now in its 11th mon...
37	Thousands of marchers took to the streets for ...
38	Hamas has described suggestions of progress on...
39	Joe Biden has said he is "optimistic" that a G...
40	Hamas has said a ceasefire plan for Gaza must ...

3.2. Term Frequency

This section focuses on the analysis of term frequency for relevant nouns identified within the 40 BBC news articles concerning the Israel-Gaza conflict. Term frequency serves as a crucial metric in understanding the prominence of specific words and concepts within a given body of text. By examining the frequency with which par-

ticular nouns appear, we can discern which terms are central to the narratives presented in these articles. A higher term frequency indicates that a word is utilized frequently throughout the news coverage, suggesting its importance within the context of the reporting. Conversely, a lower term frequency implies that a term is less prevalent, which may indicate that it plays a minor role in shaping the discourse surrounding the conflict. In our analysis, we provide a comprehensive list of the 50 most frequently occurring nouns, presented in descending order of their frequency.(Table 2) This list not only highlights the core vocabulary of the articles but also offers insights into the themes and subjects that are most salient in the coverage.

Table 2. Term frequency.

Number	Word	Frequency
1	Gaza	335
2	Israel	264
3	Hamas	212
4	people	135
5	Israeli	123
6	war	104
7	US	88
8	October	85
9	year	67
10	BBC	67
11	hostages	67
12	children	67
13	east	61
14	attack	56
15	Palestinians	53
16	day	51
17	Biden	48
18	UN	47
19	health	47
20	aid	45
21	Israels	45
22	deal	45
23	Jerusalem	44
24	Netanyahu	44
25	family	42
26	attacks	42
27	strike	42
28	ceasefire	40
29	government	37

30	conflict	35
31	Lebanon	35
32	IDF	35
33	Middle	33
34	border	33
35	Hezbollah	32
36	hostage	32
37	ministry	31
38	Iran	31
39	group	30
40	security	27
41	President	26
42	minister	25
43	civilians	24
44	forces	24
45	September	24
46	area	24
47	school	24
48	home	23
49	hospital	23
50	polio	23

It is particularly noteworthy that as illustrated in Table 2, the noun “Gaza” emerged as the most frequently mentioned term, with a total of 335 occurrences across 40 BBC news articles. This substantial frequency underscores Gaza’s significance within the reported context. In comparison, the keyword “Israel” was referenced 264 times, making it the second most frequent term. This high frequency indicates Israel’s critical role in the coverage. Additionally, “Hamas” was mentioned 212 times, ranking third in terms of frequency and further emphasizing its relevance to the discussion. Interestingly, the term “people” ranks fourth among all keywords, reflecting its centrality in the narrative. The keyword “Israeli” appears next, with 123 occurrences, positioning it as the fifth most prominent term in this analysis. The keyword “war” is also significant, ranking sixth with a total of 123 mentions, which illustrates its importance in the context of the articles examined. Furthermore, the term “Hezbollah” ranks thirty-fifth with 32 tokens, indicating that, while less prominent, it still holds a degree of relevance within the broader discussion. Other notable mentions include “ceasefire,” which was used 40 times and ranks twenty-eighth, and

“civilians,” appearing 24 times and ranking forty-third. These terms contribute to a nuanced understanding of the conflict and its implications. Additionally, “hostage” was mentioned 32 times, marking it as a significant term within the discourse.

In conclusion, the analysis clearly indicates that “Gaza” not only garnered the highest frequency with 335 tokens but also represents a pivotal focal point in the coverage of the conflict.

3.3. A Word Cloud

This section aims to analyze the word cloud generated from 40 pieces of BBC news articles. The word cloud visually represents significant keywords, with more prominent terms displayed in larger fonts and less critical ones appearing in smaller fonts. This graphical representation in **Figure 1** enables us to quickly identify essential keywords and their relative importance within the context of the news coverage. By examining the word cloud, we gain valuable insights into the central themes and topics that permeate the selected articles. This visual tool not only highlights key concepts but also provides a clearer understanding of the overarching narrative presented in these 40 pieces of BBC news.



Figure 1. A Word Cloud.

Notably, the term “Gaza” appears as the most prominent word in the word cloud, indicating that it is the most frequently used term in the analyzed articles. This prominence suggests that Gaza is a pivotal element within the discourse. Following closely, the term “Israel” ranks as the

second largest keyword. While it is less prominent than Gaza, it remains one of the most significant terms in the coverage. This highlights its essential role in the narrative surrounding the conflict. The term “Hamas” occupies the third position in terms of size, indicating its substantial presence in the articles. This frequency suggests that Hamas is also a critical keyword within the context of the news reporting. Additionally, the keyword “people” ranks fourth in size, further underscoring its importance in the discussion. This term reflects the human aspect of the conflict, emphasizing the experiences and perspectives of those affected. The noun “war” is the fifth largest term in the word cloud, suggesting that it is a central theme in the reporting. Its significant presence points to the overarching narrative of conflict that permeates the articles. Although the term “children” does not appear in the largest font size, it remains a key keyword in the discourse. Its inclusion indicates the importance of considering the impact of the conflict on younger populations, even if it does not dominate the visual representation.

In summary, the word cloud provides a valuable visual summary of the key themes and terms that shape the narrative in these BBC news articles, allowing us to grasp the significance of each keyword in understanding the broader context of the coverage.

3.4. Networks

This section is dedicated to a comprehensive analysis of the networks presented in 40 pieces of BBC news articles. **Figure 2** illustrates these networks. Network analysis serves as a vital tool for understanding the interconnections among keywords, revealing how various terms relate to one another within the context of the reporting. In the framework of network analysis, it is essential to recognize the positional significance of keywords. Core terms, which represent the most critical concepts in the discourse, are typically positioned at the center of the network. In contrast, less important keywords are situated on the periphery, indicating their subordinate role in the overall narrative. This section aims to thoroughly examine the core networks that emerge from the analysis of the selected BBC news articles. By mapping the rela-

tionships between keywords, we can gain deeper insights into the central themes that underpin the coverage, as well as the dynamics that inform public perception of the issues discussed. Through this analysis, we will identify key clusters of related terms, which will shed light on the thematic interdependencies that shape the news narrative. Understanding these networks is crucial for grasping the complexities of the issues at hand and for recognizing the broader implications of the coverage. Ultimately, this exploration of networks will enhance our understanding of how keywords interact and influence the framing of news stories, providing a richer context for interpreting the information presented in these articles.

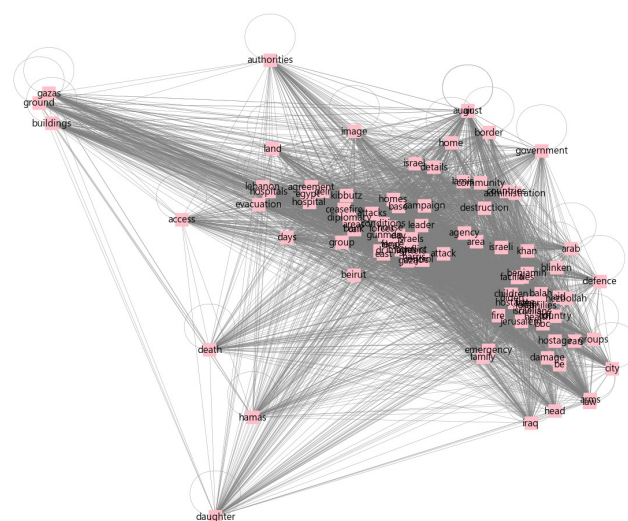


Figure 2. Networks.

It is essential to highlight that the keywords “homes,” “base,” “campaign,” “attack,” “conditions,” “group,” “leader,” and “ceasefire” prominently appear at the center of the network analysis. As previously discussed, the positioning of core words in the center indicates their significance within the narrative of the 40 pieces of BBC news. The central placement of these terms suggests that they are noteworthy and pivotal in understanding the overarching themes of the reporting. In contrast, keywords such as “Israel,” “evacuation,” “agency,” “area,” “land,” and “destruction” are situated around the center of the network. This positioning implies that, while these terms are still relevant, they are less critical than the core keywords identified earlier. Their placement indicates a secondary level of importance in the context of the news coverage.

Moreover, it is important to emphasize that terms such as “authorities,” “buildings,” “daughter,” “Iraq,” “head,” “city,” “defense,” “border,” and “government” appear on the outskirts of the network map. This positioning clearly suggests that these nouns are not pivotal when compared to the core terms, reflecting their relatively diminished role in shaping the primary narrative.

In conclusion, the analysis underscores that keywords such as “homes,” “base,” “campaign,” “attack,” “conditions,” “group,” “leader,” and “ceasefire” are indeed noteworthy in the context of the 40 pieces of BBC news. Their centrality within the network analysis indicates their crucial role in conveying the key themes and issues at play in the coverage.

3.5. A Cluster

This section is dedicated to an examination of a cluster identified within the analysis of 40 pieces of BBC news articles. In this context, a cluster refers to a group of keywords that share interrelated characteristics or thematic similarities. Utilizing Python for our analysis, we gener-

ated a variety of words that exhibit commonalities, highlighting their close relationships within the news coverage. This clustering process enables us to discern patterns and connections among terms, offering a clearer understanding of how these keywords interact and contribute to the overall narrative. It is important to note that a cluster is characterized by a degree of flexibility, distinguishing it from more rigidly defined topics. Clusters can encompass a broader range of related concepts, allowing for a more nuanced exploration of themes that emerge from the articles. To illustrate this point, we refer to **Table 3**, which presents a representation of the identified cluster. This table provides valuable insight into the interconnectedness of the keywords, facilitating a deeper understanding of the themes that dominate the reporting in these BBC news pieces. Through this analysis, we aim to shed light on the significance of these clusters in shaping the discourse and influencing public perception of the issues at hand. By understanding the relationships among these keywords, we can better appreciate the complexity of the narratives presented in the news coverage.

Table 3. Cluster.

Cluster
abbas abdeen abdel abductees abduction abductions abdul abdulrahman ability abraham absence absurdity abu abuselim abuse acceleration access accommodation accords account accounts accusations act action actionaid actions activist activities activity actors acts addition additional adel adequate adhanom adi adm administration administrations adrienne adults advantage advisers afar affairs afield afp africa aftermath agam agam age agencies agency agenda aggression agreement agricultural agriculture ahli ahmad ahmads ahmed aid aim aims air aircraft airport airspace airstrike airstrikes aisha alaa alarm albag alexander alexandra ali all allegation alleging alliance allies ally almog almost aloni alternative alumim amal ambitions ambulance america american americans ami amim amin amit amjad ammunition amnesty amount amounts amputations amusement anaesthetist analysis analyst andrei anger angrest animals anna anneliese annexe answer anti anti- semitism antoine antony antônio anything apartment apartments appeal appeals appliances application approach approaches april aqsa arab arabia arabic arabics arabs arafat arbel area areas aren argamani ariel ariev arm arms army arnon arrangements arrogance arsenal arthur asdaa ash ashdod ashraf assad assassination assault assaults assembly assessment assistance association assumptions atia atid atmosphere atrocities attack attacks attempt attempts attorney attrition audience audiences august aunt austin austins australia authorities authority avi- natan avital aviv aviva avshalom awali aware axios axis aya ayalon ayatollah aza babies baby back backdrop background backlash badran baffling bags baher balah balfour ballot bank banners bar barah barely bargain barriers base baseball bases bashar basis battle battleground battles bbc bbcs be beach beaches beanie beans beard beauty bedfellow bedouin bedouins beeri beginning behaviour beidoun beirut beit belief bell belongings ben benjamin bennett benny bereaved berger berman between bezalel bias bibas bicycle Biden bidens bilal bill bins bipin birth bit Blackburn blame blames blanket blast blinken blinkens block blockade blogger blood blow blows blue blueprint board boats bodies body bomb bombardment bombardments bombers bombing bombings bombs bonkowsky border borders born both boundaries bow- en boxes boy boyfriend boynizam brain brand breakdown breaker breakneck breakthrough breath breathing brick bridging brief brigade Britain British broadcaster broadcasters Brockenhurst broke brother brothers brown brunt brushwood brutality btselem budgeting building buildings bullet bunker burkman burn burns burnt bus buses business businesses buster cabinet cafes Cairo calculation California call calls calories cameras camp campaign campaigns camps candidate candle cannot cap capabilities capacity capital captives captivity captor cap- ture captures car carcass cardamom care career carer carers carmel carmel carrier cars case cases cash casualties casualty cat catastrophes category catherine cbs ceasefire ceasefires census center central centre centres century ceo ceremonies ceremony chain challenge challeng- es chamber chance chancellor chances channel channels chanting chants chapter charge charities charity charredreuters chart checkpoint checkpoints checks chen cheney cheneys chenjonathan Chicago chicken chief chiefs child children childrens childs chin china choice chris christian chronicle cia circumstances cities citizen citizens citizenship city civilians clad claim claiming claims clara clarifications classifi- cation clearance cleveland clinic clinics clinton clip clock close closer closest cloth clothes clothing cloud cluster clusters co coalition coast coastline coffee cogat cohen cold collapse colleagues collective colorado combat combatants command commander commanders

commemorations commentators commerce commission commitment committee commons communication communications communities community complain complications composite compound compounds comprehension compromise compromises concern concerns concessions condition conditions conduct confirmation conflict conflictgetty conflicts consequences considers consignment construction constructive contact contagion container containers contamination content contents context contingent contract contraction contracts contradiction contrast control convention conversions conviction cooking coons coop coordination coordinator cop core cornel corridor corridors cost council counter counterparts countries country couple course court courts cover coverage cracks cranes cream creation credibility crescent crickx crime crimes criminal crisis criticism critics cross crossing crossings crowd crowds cumbersome cummings cunio cuny currency customs cycle dad dafna daher dahieh dairy damage damari dan dana danger daniel daniela daniela danino dare dark darwish data daughter daughters david dawn day daylight days deadliest deal dearborn death deaths debris dec decade decades december decision decisions declaration deeply defeat defence defense deir dekel delay delegation deliberate deliveries delivery demand demands demise democrat democrats demolition demolitions demonstration demonstrations demonstrator demonstrators department departments depend deployment deraawi dermer desalination descendants descent described description desert designate desire desperate destruction details detainees deterioration determination detroit devastation development devices diagnosis dialogue dick didnt difference differences difficulty dignity dilemma diplomacy diplomat diplomatic direction director dirty disasters discharge discrimination discussions disease diseases disfigurement displaced displacement disrepair disruption distance distinct distraction distribution district districts distrust ditza division divisions dnc doctor doctors dodds doesn't doha dollars don donald donations donnelly donnellypa dont door doron dose doses dozens dr drab drainage drama drawing dream dreams dress drinking drive drivers drone drop dust duty earlier earth earthquake east eastern economy edan eden edge editor education efe effort efforts efrat eggs egypt egyptian egyptians ehud eitan el election elections elgarat eli eliya elkadi ema embargo embassy emergence emergency emilia emily emir emmanuel empire end endeavour endorsement enemies enemy engel enraged entire entry epa equipment equivalent erez eri escalation essentially establishment estimates eu europe european evacuation evacuations eve evening event events eventually every everyone everything evidence evil examination example examples exasperation except exchange exchanges executive existence expansion experience expert experts explosions exposure extension extent extermination extremist extremists eyes eyewitness facades face facilities facility fact facts fadl faeces failure falafel falah families family familys famine far farewell farhan farmer fatah fate father fault favour fawzia faye fears february feel feelings feels feet fence fernando ferocity ferris festival festivalreuters fever field fields fight fighter fighters fighting figure figures filipina film filreuters finance findings fines fire fireballs fires firing flag flags flames flash fleeing flight flooding flow focus foetus follow food footage football force forces foreground foregroundpa foreign foreigners forerunner former forms forty forum foundation founder four framework france francis freed freeing freelancer freezer french friction friday fridge friend friends front fronts fuel fun functions fund fundraising funds future gabriela gadi gali gallant gallants game gangs gantz gaps gas gat gathering gauntlet gazafamily gazan gazans gazas gaze gen general generals generation generations genocide george german germany gerstner getty ghadir ghebreyesus gifts gilboa gilo girl girls gisha glance glasses go goal god goes golan golani goldberg good goodbye goren government governments governorates graduate granddaughter grandmother grandson graphic greater green grenades grey grief gritzewsky ground groundmandy group groups growth guard guards guide guilt gunfight gunfire gunmen gutters guys gvir haganah hagari haifa hair hajji half halford halimy hallway hamade hamal hamas hamass hamid hammoud hamouda hampshire hamza hand handbag handcuffs handful handling handout handouts hands haneen hangs hanyeh hanoun har haran harder harm harris harrison hassan hat haven hayya head headquarters heads headscarf heal health healthcare heard heart hebrew heights hell help helplessness her hersh herzog hezbollah hezbollahs hiba hid hider highway hila hilafamily hind hints hips histadrut history hit hitting holding hole holes holit holocaust home home homeland homelands homes horn horror hospital hospitals hostage hostages hostilities hotel hour hours house houses housing houthis how hse huldai hum human humanitarian humanity humiliation humility hundreds hunger hurdle husband hussam hussein husseini hygiene ibrahim icons idan idea identification ideology idf idfs ilana illegal illusion illusions illustration im image imagery images iman immunisation immunisations immunities impact implementation implication implications imports improvements incident incidents include increase increases incursion incursions independence indications indirect indiscriminate individuals industry infantry infection infections inflation influxes information infrastructure inhabitants injured injuries inland inquiry inset insists instability instagram institute institutions insurance intelligence intent intentions international intervention investigation introduction invasion investigations involvement ipc iplayer ir iran iranians irans iraq iraqi irc ireland irina irish iron is isaac islamic islamist ismael ismail isn isolation israa israeli israelis israel's issue issues itamar itay items itv itzik ive jabal jabalia jails jan january jaundice jeep jemin jeremy jerusalem jets jewish jews jidyan jihad jihads jill jisoor job joe joint jonathan jordan joshi joshua journalist journalists journey juda judyan july june just justice kaid kalderon kamala karim karina karyna katz keir keith kerem key kfar kfar kg khaled khalid khalil khamenei khan kibbutz kibbutzim kids killing killings kilograms kilometre kind king kissufim kitchen kitchenreuters km knowledge kozlov labor labour labours lack ladle lagoon lama lamis lammy land landscape language lap lapid latest law lawlessness lawmaker lawn laws lawyer lay layers lazimi lead leader leaders leadership league leaks lebanese lebanon left leg legion legs leimberg lemus lena lens let letter level levels leverage levy liability liberation lies lieutenants lifeline lifshitz likelihood limbs limit limits lina lincoln line lines lingers lip liri lishay list lives living ll lloyd lobanov location locations logistics london londoners lookout looters lorries loss losses lot louis louise loyalist lungs lutfiyah lutfiyahs maayan machinery macron magen mahmoud maisa majad major majority makeshift malka maman mamode man management manager manchester mandate mandy mann manner mansour many map maps march marchers margalit margin markings marks

In our analysis, we identified a substantial number of words that share common characteristics; however, due to space constraints, we have chosen to present only a portion of the cluster in this discussion. As illustrated in Table 3, these words frequently co-occurred in the 40 pieces of BBC news, suggesting that they collectively contribute

to a coherent narrative. These keywords can be seen as forming a cohesive group, reflecting their interrelated nature within the context of the reporting. This interconnection allows us to understand how these terms work together to convey specific themes and ideas, thereby earning the designation of a “cluster.” By examining this

cluster, we can glean insights into the overarching messages and implications presented in the news coverage. The relationships among these keywords not only enhance our comprehension of the individual terms but also illuminate the broader narrative constructed by the articles. Ultimately, the identification of this cluster underscores the importance of contextual analysis in news reporting. Understanding how these terms interact within a cohesive framework allows for a richer interpretation of the issues being discussed, revealing the complexity and depth of the narratives presented in the selected BBC news pieces.

3.6. Topics

This section is dedicated to an exploration of the main topics identified within the 40 pieces of BBC news articles. By analyzing these articles, we aim to discern which topics are most likely to emerge as central themes in the coverage. Specifically, this analysis involves assessing the probability of certain keywords becoming predominant topics. Understanding these probabilities allows us to gauge the significance of various themes and how they resonate within the broader context of news reporting. In **Table 4**, we present 16 keywords identified as main topics, along with their associated probabilities. This quantitative assessment not only highlights the relevance of each topic but also provides a framework for understanding the relationships among them. By examining these main topics and their probabilities, we can gain valuable insights into the issues that are most salient in public discourse. This understanding is crucial for interpreting the narratives constructed by the BBC news articles and for recognizing the prevailing trends that shape the media landscape. In summary, this section aims to provide a comprehensive overview of the primary topics emerging from the analysis of the selected news pieces, thereby contributing to a deeper understanding of the thematic focus within BBC's reporting.

Table 4. Topics.

Number	Word	Probability
1	Gaza	0.105
2	Israel	0.084
3	Hamas	0.074

4	children	0.022
5	day	0.017
6	attack	0.015
7	Palestinians	0.014
8	Lebanon	0.013
9	UN	0.011
10	Hezbollah	0.008
11	BBC	0.007
12	hostages	0.007
13	Lamis	0.006
14	military	0.005
15	restrictions	0.005
16	UK	0.004

As indicated in **Table 4**, the term “Gaza” stands out as the most significant topic in our analysis, exhibiting a probability of 10.5% being identified as a main topic. This notable prominence underscores the central role that Gaza plays in contemporary discussions and conflicts in the region. Following Gaza, the term “Israel” ranks second with a probability of 8.4%. This ranking indicates the intertwined nature of the narratives surrounding both Gaza and Israel, suggesting that discussions often revolve around these two entities in relation to one another. The significance of Israel in this context cannot be overstated, as it remains a pivotal player. In third place is “Hamas,” which carries a probability of 7.4%. This reflects the organization’s influential position within the discourse surrounding Gaza and Israel. The prominence of Hamas in discussions highlights the complexities of the conflict and the various actors involved, further emphasizing the challenges to peace and stability in the region. The term “children” ranks fourth, with a probability of 2.2%. This statistic is particularly poignant, as it draws attention to the humanitarian aspects of the conflict. The mention of children in this context serves as a reminder of the impact that ongoing violence has on the youngest and most vulnerable populations. In sixth place is the term “attack,” which has a probability of 1.5%. This term encapsulates the violent incidents that frequently dominate media coverage and public discourse, illustrating the urgent security concerns that permeate the region. Additionally, “Hezbollah” ranks tenth, with a probability of 0.8%. The inclusion of Hezbollah in this analysis points to the broader regional dynam-

ics and the multiple groups that influence the situation in Gaza and Israel. Finally, the term “hostages” occupies the eleventh position, with a probability of 0.7%. This term underscores the human cost of the ongoing conflict and highlights the distressing realities faced by those caught in the crossfire. In summary, our analysis reveals that “Gaza” holds the highest probability of being a main topic, closely followed by “Israel.” This interconnectedness of terms underscores the complexity of the situation and the necessity for a nuanced understanding of the various factors at play.

3.7. The Similarity between Words

In this section, we delve into the concept of similarity between words, focusing on the relevance of various word pairs within the analyzed corpus. The term “similarity” encompasses the degree to which two or more words are related in meaning, context, or usage, and it plays a crucial role in understanding the nuances of language in news articles. By examining word pairs (**Table 5**), we can uncover patterns of association that provide insights into how specific terms interact with one another. This analysis allows us to explore the thematic connections that exist within the discourse, highlighting the relationships that inform the overarching narratives presented in the BBC news coverage. To assess the similarity between words, we utilize various linguistic and computational techniques, which enable us to quantify these relationships. Through this methodology, we can determine which words frequently co-occur, indicating a strong semantic or contextual linkage. Understanding the similarity between words is particularly important in the context of news reporting, as it helps to illuminate the underlying themes and issues being addressed. By identifying clusters of related terms, we can better comprehend the complexity of the narratives presented and the dynamics that shape public perception.

In summary, this exploration of word similarity provides a valuable framework for analyzing the interconnections among terms in news articles. By shedding light on these relationships, we can deepen our understanding of the language used in reporting and the implications it carries for the interpretation of current events.

Table 5. The Similarity between Words.

Number	Word	Word	Similarity
1	Hamas	Gaza	0.9999499
2	Hamas	BBC	0.9999496
3	Hamas	Israeli	0.9999461
4	Hamas	Israel	0.9999384
5	Hamas	war	0.9999349
6	Hamas	US	0.9999327
7	Hamas	people	0.9999256

It is noteworthy that as illustrated in **Table 5**, the similarity between the terms “Hamas” and “Gaza” is the highest among all relevant keywords, with a remarkable similarity score of 99.99499%. This indicates a profound contextual relationship between these two terms in the 40 pieces of BBC news analyzed. Following closely, the term “BBC” has the second highest similarity score of 99.99496% in relation to “Gaza.” This suggests that the context in which these terms are used is closely intertwined, reinforcing the significance of their association in the reporting. Additionally, it is significant to observe that the similarity between “Hamas” and the “US” ranks sixth, with a score of 99.99327%. This high degree of similarity indicates that discussions involving Hamas often reference the United States, highlighting the geopolitical dimensions presented in the coverage. Furthermore, the similarity between “war” and “Hamas” ranks fifth, with a score of 99.99349%. This emphasizes the interconnectedness of these concepts within the context of the articles, suggesting that narratives around conflict frequently involve both terms. In conclusion, the analysis clearly demonstrates that the similarity between “Hamas” and “Gaza” is the most pronounced among the keywords examined, indicating their central role in the discourse surrounding these news articles. Now, let us turn our attention to **Table 6**, which provides a detailed representation of these similarity scores and further illustrates the relationships among the terms discussed.

Table 6. The similarity between words.

Number	Word	Word	Similarity
1	Gaza	BBC	0.9999550
2	Gaza	Hamas	0.9999500
3	Gaza	Israel	0.9999494

4	Gaza	US	0.9999455
5	Gaza	children	0.9999380
6	Gaza	IDF	0.9999380
7	Gaza	family	0.9999347

It is essential to highlight that the similarity between “Gaza” and “BBC” holds the highest proportion among the relevant keywords examined. This significant similarity indicates a close relationship in the contexts presented across the 40 pieces of BBC news analyzed. Following closely, the term “Hamas” ranks as the second most similar term to “Gaza,” with a striking similarity score of 99.995%. This high degree of correlation implies that these two terms are intimately related within the reporting, reflecting the interconnected narratives often presented in the news. Additionally, it is noteworthy that the similarity between “Gaza” and “Israel” stands at 99.99494%, marking it as the third highest. This indicates a significant contextual overlap between these two terms, suggesting that discussions surrounding Gaza frequently involve references to Israel. Moreover, the similarity between “Gaza” and “family” ranks seventh, with a score of 99.99347%. This suggests a notable interrelationship, emphasizing the human aspect of the discourse in the context of the articles. In conclusion, the analysis clearly demonstrates that the similarity between “Gaza” and “BBC” is the most pronounced among the keywords explored. This finding underscores the importance of understanding of how these terms are interwoven in the narratives presented in the news coverage. For the corpus analysis of words, see [2,7–20]. Tognini-Bonelli ^[9] explores the practical applications of corpus linguistics, emphasizing its role in linguistic analysis. Tognini-Bonelli’s second work further delves into corpus linguistics, applying it to various linguistic phenomena.

4. Discussion

This study examines the portrayal of Israel and Hamas in 40 BBC News articles, utilizing Python for text analysis through six distinct methods. BBC News does not take a stance in favor of either Israel or Hamas; rather, it maintains a neutral position.

Key findings reveal that “Gaza” is the most frequently mentioned term, appearing 335 times, which underscores its central role in the coverage. “Israel” follows with 264 mentions, indicating its significance, while “Hamas” is referenced 212 times, further emphasizing its relevance in the narrative. The term “people” ranks fourth, reflecting the human impact of the conflict. The analysis also highlights other notable keywords, such as “homes,” “campaign,” “attack,” “conditions,” “group,” “leader,” and “ceasefire,” which are integral to understanding the themes presented in the articles. The prominence of “Gaza” is quantified, showing a probability of 10.5% of being identified as a main topic. In comparison, “Israel” holds an 8.4% probability, and “Hamas” comes in at 7.4%. This suggests an intertwined relationship in the narratives surrounding these terms. Moreover, the study identifies a remarkable similarity score between “Hamas” and “Gaza” at 99.99499%, indicating a strong contextual connection. The term “BBC” also displays a high similarity to “Gaza,” reinforcing the context in which these terms are reported. Following closely, “Hamas” maintains a similarity of 99.995% with “Gaza,” and “Gaza” and “Israel” share a similarity score of 99.99494%. These findings emphasize the interconnectedness of the terms and their roles in shaping public perception and discourse about the conflict. Overall, the analysis concludes that BBC News maintains a neutral stance, refraining from favoring either Israel or Hamas. This impartiality is supported by various analytical approaches, including term frequency, word cloud, network analysis, topic analysis, and similarity analysis, demonstrating the complexity of the narratives involved in contemporary geopolitical issues. This study highlights the significance of language in influencing public understanding and framing discussions around the Israel-Hamas conflict.

Future research could explore the nuanced use of language in news reporting by conducting a detailed linguistic analysis of the terminology surrounding key terms like “Gaza,” “Israel,” and “Hamas.” A comparative study of language use in different media outlets could help identify patterns of bias or neutrality. Additionally, examining how emotional language influences reader perceptions could provide insights into the effects of framing on public

opinion. Investigating the role of metaphors and imagery in shaping narratives would deepen our understanding of discourse. Finally, longitudinal studies could assess how language evolves over time in relation to major events in the region.

5. Conclusions

In summary, this paper has provided an in-depth analysis of 40 pieces of BBC news coverage from 2024 concerning the Israel-Gaza war. Our findings indicate that the noun “Gaza” achieved the highest frequency, with a total of 335 tokens, positioning it as the most significant term in the context of the reporting. This prominence is further evidenced by its status as the largest word in the word cloud, suggesting its pivotal role in the narrative. Additionally, the term “Israel” ranks as the second most prominent keyword. While it is less central than “Gaza,” it remains a crucial term within the discourse, highlighting its importance in discussions surrounding the conflict. Throughout this analysis, we identified several noteworthy keywords, including “homes,” “base,” “campaign,” “attack,” “conditions,” “group,” “leader,” and “ceasefire.” These terms play an essential role in shaping the narrative presented in the selected news articles. We also determined that “Gaza” has the highest probability of being identified as a main topic, with a likelihood of 10.5%. It is followed closely by “Israel,” which has an 8.4% probability of serving as a main topic. This quantitative assessment reinforces the significance of these terms in the context of the news coverage. Furthermore, our analysis highlighted the similarity between “Gaza” and “BBC” as the highest among the keywords examined, indicating a close relationship in the reporting. The similarity between “ Hamas” and “Gaza” ranks second, further illustrating the interconnectedness of these terms within the narrative. The insights derived from this analysis contribute to a deeper understanding of the key themes. These findings underscore the importance of language in shaping public perception and discourse surrounding complex geopolitical issues. In conclusion, BBC News does not take a stance in favor of either Israel or Hamas; rather, it maintains a neutral position. This impartiality is effectively illustrated through various analytical

methods, including term frequency analysis, word cloud analysis, network analysis, topic analysis, and similarity analysis.

Appendix A

In the following, we undertake a comprehensive analysis of 15 CNN news articles using two key methods: term frequency analysis and word cloud analysis. These analytical approaches will allow us to explore and compare the language used in CNN articles with that found in BBC news articles, shedding light on potential differences in tone, word choice, and framing. By focusing on specific linguistic patterns, we aim to identify whether the language in these articles may reveal a particular stance towards the ongoing conflict between Israel and Hamas. As outlined previously, our analysis is based solely on nouns. To ensure the accuracy and focus of our study, we first removed all adjectives, adverbs, to-infinitives, gerunds, relative pronouns, and verbs. This narrowing of the data set to nouns is crucial for both term frequency analysis and word cloud analysis, as it allows us to concentrate on the key substantive terms that appear most frequently in the articles. Nouns, being the primary carriers of meaning in a text, provide the most direct insights into the core topics and themes under discussion. Term frequency analysis enables us to determine how often particular words appear within the text, providing a quantitative measure of their prominence. This analysis is particularly useful in identifying which terms are most central to the narrative of each article. For example, if certain words related to Israel, Hamas, or the broader conflict appear more frequently, it may indicate a particular focus or emphasis in the reporting. Furthermore, by comparing the frequency of relevant terms across multiple articles, we can assess whether there is any noticeable bias or divergence in the language used by CNN, especially when compared to other media outlets like BBC. In parallel, word cloud analysis offers a visual representation of the most frequent terms within the articles. The larger the size of a word in the cloud, the more frequently it appears in the text. This visual approach allows us to quickly identify the words that dominate the discourse, providing an intuitive sense of what topics or

issues are being highlighted. Word clouds can also reveal subtler linguistic trends, such as whether certain terms are used in a positive or negative context, or whether particular concepts are being prioritized over others. Together, these two methods—term frequency and word cloud analysis—allow us to predict whether the 15 CNN articles exhibit a linguistic bias or alignment with either Israel or Hamas. By examining which words are emphasized and how they are presented, we can infer the possible stance taken by the articles, whether explicit or subtle. The underlying assumption is that the language used in these reports plays a critical role in shaping public perception and framing the conflict. The following section presents the first sentence of each of the 15 CNN articles we analyzed. These introductory sentences serve as a starting point for our examination, offering a glimpse into the themes and language that dominate the rest of the article. By analyzing these samples in conjunction with the term frequency and word cloud data, we aim to build a more nuanced understanding of how language influences the portrayal of the Israel-Hamas conflict in CNN’s coverage.

Table A1. Sample of 15 CNN articles.

Number	Sample
1	In the aftermath of an Israeli airstrike that ...
2	An Israeli airstrike on a multistory building ...
3	US Defense Secretary Lloyd Austin spoke with h...
4	A 40-year-old father of four, Eliran Mizrahi d...
5	Israel’s parliament defied international criti...
6	An Israeli strike killed several people at the...
7	More than 40,000 Palestinians have been killed...
8	The head of the United Nations’ Palestinian re...
9	Mohammad al-Dalou whimpered as he mourned the ...
10	Lebanon marked the most fatalities in a single...
11	The US, Qatar and Egypt called Saturday on Isr...
12	Thousands are fleeing the southern city of Raf...
13	US President Joe Biden offered one of his shar...
14	A UNICEF official said Monday that warnings ab...
15	US President Joe Biden acknowledged the “gut-w...

After extracting the nouns from the texts, we employed Python to calculate the term frequency of the relevant words. To optimize for clarity and brevity, we selected the 40 most frequently occurring keywords, presented in descending order of frequency:

Table A2. Term frequency.

Number	Word	Frequency
1	Gaza	110
2	Israel	59
3	Israeli	58
4	Hamas	53
5	people	50
6	US	35
7	war	29
8	CNN	28
9	ceasefire	25
10	children	24
11	hospital	24
12	defense	20
13	aid	19
14	Biden	19
15	IDF	19
16	health	18
17	hostages	18
18	deal	17
19	forces	17
20	soldiers	17
21	civilians	16
22	UN	16
23	minister	14
24	talks	14
25	strike	13
26	conflict	12
27	family	10
28	government	10
29	agency	9
30	attack	9
31	death	9
32	home	9
33	leader	9
34	media	8
35	response	8
36	mediators	7
37	refugee	7
38	release	7
39	situation	7
40	negotiations	6

It is important to highlight that the term “Gaza” emerged as the most frequent and prominent keyword across both sets of articles, with a total of 110 occurrences.

es. This indicates its central role in the discourse of both the 40 BBC and 15 CNN news articles analyzed. The frequency of this term underscores its significance in the coverage of the Israel-Hamas conflict, suggesting that “Gaza” was the most noteworthy term in the articles studied. Following “Gaza,” the term “Israel” ranked second, with 59 occurrences. Notably, both the BBC and CNN articles exhibited a similar pattern in their use of this term, further reinforcing its centrality to the reporting. The proximity of these two terms—“Gaza” and “Israel”—in terms of frequency suggests a clear thematic focus on these two entities within the articles. Additionally, the term “Israeli” ranked third, with 58 occurrences. Interestingly, while it was the third most frequent term in the CNN articles, it held only the fifth position in the BBC articles. This difference suggests that while both news outlets focus heavily on Israel, the positioning of “Israeli” in the frequency ranking varies slightly between the two sources. In the BBC articles, the term “Israeli” ranked fifth, with 56 occurrences, and in CNN articles, it was positioned more prominently, reflecting a slightly different emphasis in coverage. The term “Hamas” ranked fourth in frequency, with 53 occurrences in the CNN articles, suggesting a similar pattern in its usage across both sets of news articles. In the 40 BBC articles, “Hamas” ranked third, with 212 occurrences, further reinforcing its importance in the coverage of the conflict. It is also noteworthy that the term “Hamas” was closely followed by the term “people” in both the BBC and CNN articles, with “people” ranking fourth in BBC articles and third in CNN articles. This consistent ranking of “Gaza,” “Israel,” “Israeli,” “Hamas,” and “people” across both sets of articles indicates that the overall pattern of term frequency remains largely similar in both the BBC and CNN coverage. Specifically, “Gaza” emerged as the most frequently used term, followed by “Israel,” “Israeli,” “Hamas,” and “people.” These findings suggest that both BBC and CNN share a comparable focus on the key themes surrounding the conflict, without showing a marked difference in their choice of terminology. Most notably, the frequency of the terms “Israel” and “Hamas” in the CNN articles did not show a significant disparity, with both terms being used almost equally. This lack of variation in the frequencies of these terms suggests

that CNN's coverage maintains a neutral stance, similar to BBC's approach. The even-handed usage of these terms in both sets of articles indicates that neither side of the conflict is given a disproportionate amount of attention, further supporting the conclusion that both news outlets strive for neutrality in their reporting. With these findings in mind, we now turn our attention to the word cloud analysis, which provides a more visual representation of the term frequencies and may offer additional insights into the linguistic patterns in the articles.

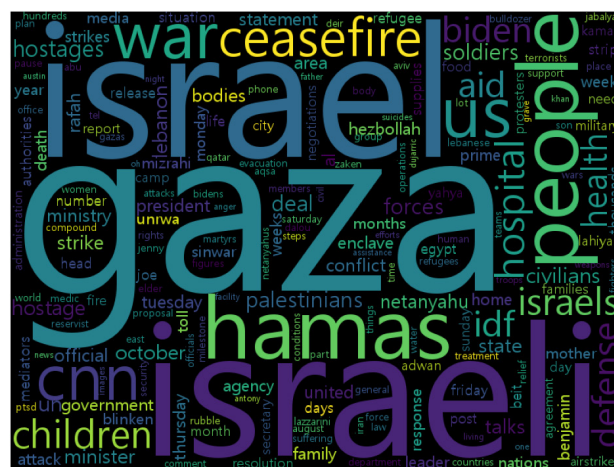


Figure A1. A Word Cloud.

As illustrated in **Figure A1**, the term “Gaza” appears with the largest font size, signifying its prominence among the key terms analyzed. This visual representation underscores “Gaza” as the most noteworthy and central term in the context of both the CNN and BBC news articles. In the word cloud for the 40 BBC news articles, the term “Gaza” similarly occupies the most significant position, confirming its importance in the overall discourse of the conflict. Interestingly, the term “Israel” ranks second in prominence, with a font size that places it just below “Gaza” in both the CNN and BBC word clouds. This pattern suggests that “Israel” is the second most significant term, reflecting its centrality in the coverage of the conflict. The consistent ranking of “Israel” as the second most prominent term across both media outlets further supports its pivotal role in shaping the narrative. It is also worth noting that while the term “Hamas” appears in the word cloud, it is consistently smaller in size than “Israel,” though the difference in size between these two terms is not substantial. In both the

CNN and BBC word clouds, “*Hamas*” follows closely behind “*Israel*” in frequency but is generally less prominent in terms of font size. This similarity in ranking between the two outlets suggests that “*Hamas*” is recognized as an important subject in the coverage, albeit with slightly less emphasis compared to “*Israel*.” The term “*people*” also appears in the word cloud but with a smaller font size than “*Hamas*.” This indicates that “*people*” is a less prominent term in the overall coverage, suggesting that it plays a more peripheral role in the articles analyzed. These visual and quantitative analyses suggest that both CNN and BBC exhibit similar patterns with respect to the size and prominence of major key terms. Specifically, “*Gaza*” stands out as the most significant term, followed by “*Israel*,” “*Hamas*,” and “*people*.” This consistent pattern across both sets of articles implies that neither outlet places overwhelming emphasis on one side of the conflict over the other. Therefore, it is reasonable to conclude that both CNN and BBC maintain a relatively neutral stance, avoiding a clear bias in favor of either Israel or Hamas. This neutrality is reflected in the balanced representation of key terms in the word clouds, where no single side dominates the discourse.

Funding

This work was supported by the 2024 Far East University Research Grant (FEU2024S01).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

This study employs a systematic approach to analyze 40 BBC news articles and 15 CNN news articles pertaining to the Israel-Gaza war broadcasted in 2024. Each article was obtained from the BBC’s online archives and the CNN’s online archives. These articles are still available in Google. For more details, refer to the samples of each arti-

cle (Section 3.1) and the Appendix A of this paper.

Acknowledgments

My thanks are owed to Far East University in South Korea for the Far East University Research Grant.

Conflicts of Interest

The author declares no conflict of interest.

References

- [1] Gee, J.P., 2014. *An Introduction to Discourse Analysis: Theory and Method*. Routledge: New York, USA. pp. 1-250.
- [2] Biber, D., Conrad, S., Reppen, R., 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press: Cambridge, UK. pp. 1-400.
- [3] Halliday, M.A.K., Matthiessen, C., 2014. *An Introduction to Functional Grammar*. Routledge: London, UK. pp. 1-500.
- [4] Swales, J.M., Feak, C.B., 2004. *Academic Writing for Graduate Students: Essential Tasks and Skills*. University of Michigan Press: Ann Arbor, USA. pp. 1-300.
- [5] Van Dijk, T.A., 1997. *Discourse as Structure and Process*. Sage Publications: Thousand Oaks, USA. pp. 1-400.
- [6] Sampson, G., 2001. *Empirical Linguistics*. Continuum: London, UK. pp. 1-250.
- [7] Baker, P., 2006. *Using Corpora in Discourse Analysis*. London: Continuum: London, UK. pp. 1-250.
- [8] McEnery, T., Wilson, A., 2001. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press: Edinburgh, UK. pp. 1-350.
- [9] Tognini-Bonelli, E., 2001. *Corpus Linguistics at Work*. John Benjamins Publishing: Amsterdam, Netherlands. pp. 1-250.
- [10] Sinclair, J., 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press: Oxford, UK. pp. 1-200.
- [11] Gries, S. T., 2009. Quantitative Approaches in Usage-based Linguistics. In: *Handbook of Cognitive Linguistics*. New York: Routledge: New York, USA. pp. 178-195.
- [12] McEnery, T., 2011. *Corpus Linguistics: Method, theory and practice*. Palgrave Macmillan: London, UK. pp. 1-300.
- [13] Baker, P., McEnery, T., 2005. A Corpus-Based Study of Euphemism and Dysphemism. *Journal of Prag-*

- matcs. 37(6), 955-967.
- [14] Lott, D., 2009. The Use of Corpora in the Teaching of English for Academic Purposes. *English for Specific Purposes*. 28(2), 119-130.
- [15] Charles, M., 2011. A Corpus-Based Study of Academic Writing. *Journal of English for Academic Purposes*. 10(3), 213-226.
- [16] McCallum, A., 2005. MALLET: A Machine Learning for Language Toolkit. Available from: <http://mallet.cs.umass.edu> (cited 2 December 2, 2024).
- [17] Taboada, M., Grieve, J., 2004. Analyzing Sentiment in Text. *Computational Linguistics*. 30(3), 259-270.
- [18] Evison, J., Turner, K., 2014. Using Corpora for Text Analysis in Research. *Language Resources and Evaluation*, 48(1). 69-85.
- [19] Rayson, P., Garside, R., 2000. Comparing Corpora Using Frequency Profiling. *Proceedings of the Workshop on Linguistic Data Consortium*; May 30-June 1, 2000; Philadelphia, USA. pp. 1-15.
- [20] Baroni, M., Bernardini, S., 2006. Bootstrapping the WordNet with a Corpus-Based Approach. *Proceedings of the International Conference on Language Resources and Evaluation*; May 22-28, 2006; Genoa, Italy. pp. 1-10.