1 gram Analysis.

This analysis investigates the presence and prominence of violence related language in Al-Jazeera's coverage of Gaza by using a distant reading method. By analyzing unigram frequency data over time, we argue that violent terms such as kill, bomb, strike, and casualty exhibit noticeable spikes during certain years likely corresponding to escalations in conflict. The resulting visualization not only highlights the intensity of media focus on violence but also reveals how different terms gain or lose prominence across time. This reveals the media’s shifting narrative emphasis and supports a critical reading of how language shapes public perception of war and suffering in Gaza.

Step-by-Step Methodology

1. Data Loading and Exploration

We began by importing the 1-gram dataset using pandas, which contained:

* Unigrams (single words)
* Year of appearance
* Total frequency (count-sum) per year

I explored the data to:

* Identify the most common words
* Investigate key unigrams like "Gaza" and "war"
* Observe which words are repeated most often across years

2. Keyword Selection and Filtering

I manually defined a list of violence-related base keywords:  
["attack", "bomb", "kill", "strike", "casualty"]

Using str.contains() and regular expressions, I filtered the dataframe to only include rows where the unigram matched any of these base keywords or their morphological variants (like “killed”, “bombing”).

3. Categorization Using a Helper Function

To organize our data by theme, I wrote a function find\_keyword() that:

* Checked each unigram
* Assigned it to one of the five base keywords if matched

This allowed me to categorize related terms under one umbrella (e.g., “killed”, “killing”, “kills” all grouped under “kill”), simplifying the visualization.

4. Aggregating Data by Year and Keyword

I grouped the filtered data by year and keyword using groupby(), then summed the total mentions (count-sum). This created a clean table showing how often each violence-related keyword appeared per year.

5. Visualization with Stacked Bar Chart

I used Plotly Express to plot a stacked bar chart:

* X-axis: Year
* Y-axis: Total mentions of violent terms
* Color: Each bar segment represented a keyword

This form allowed me to see both the total volume of violence-related discourse per year and the composition of that discourse (i.e., which types of violent language were most prominent).

6. Exporting the Graph

The chart was saved as an interactive HTML file (Abdus salam.html), allowing viewers to hover, zoom, and explore yearly spikes.

Analysis of the Graph

The graph shows clear spikes in the frequency of violence-related terms in specific years—particularly noticeable in the usage of “kill”, “bomb”, and “attack”. These spikes likely correspond to major escalations in the Gaza conflict, reflecting media responses to real-world violence. For example, a large rise in "kill" may indicate a period of intense reporting on casualties, while "strike" or "bomb" may dominate when airstrikes or military operations are the focus.

The use of a stacked bar chart is significant because it not only shows the total discourse volume but also reveals shifts in emphasis. In some years, “attack” dominates the stack, while in others, “casualty” rises, suggesting a narrative move from reporting military actions to reporting human loss.

This project demonstrates how distant reading by focusing on frequency and pattern rather than close interpretation can offer powerful insights into media framing, violence discourse, and temporal trends. Rather than reading every article individually, this method allowed me to trace how language changes over time and how specific themes (like violence) recur in media representations of Gaza.

By filtering, grouping, and visualizing violence-related language in the Gaza corpus, this analysis supports the argument that media coverage often pivots around cycles of conflict. The dominance of certain words in specific years aligns with real-world events and reveals how the language of violence is not constant but instead shifts in intensity and focus. This method offers a scalable way to criticize narrative framing in war reporting and emphasizes the power of words in shaping collective memory.