

# Analysis of Kamala Harris News Coverage in the Top Mainstream U.S. News Outlets Following the 2024 U.S. Election

Jacqueline Wu (260394630 - McGill University), Erin Wyndham (260984817 - McGill University),  
Greta Ru-Mei Zu (261099681 - McGill University)

[yiru.wu@mail.mcgill.ca](mailto:yiru.wu@mail.mcgill.ca), [erin.wyndham@mail.mcgill.ca](mailto:erin.wyndham@mail.mcgill.ca), [greta.zu@mail.mcgill.ca](mailto:greta.zu@mail.mcgill.ca)

## Introduction

The U.S. election is always a major event in North American news; this year's is no exception. The most recent election from November 2024 featured Donald Trump representing the Republican party and Kamala Harris representing the Democratic Party as the two main candidates running for the office of the president of the United States of America.

This event generated a lot of media coverage, in particular, there has been significant interest toward Kamala Harris, not only because she was a fairly new candidate after Biden's recent decision to withdraw from the election race, but also because she stood high chances of becoming the first female president.

To better understand what the public is seeing in the news regarding the election, a media company hired us with a particular interest in the portrayal of Kamala Harris in American news since the election, a pivotal moment for her public image. To fulfill these requirements, our report analyzes the different positive and negative topics concerning Kamala Harris in the top mainstream U.S. news outlets by online readership following the 2024 U.S. election. Our stakeholders include the media company, members of the democratic party, as well as supporters of the democratic party who may be interested in how the media is portraying their representative. Oftentimes, the general public also bases their personal opinions on the general consensus suggested by news outlets, so these results would also impact them.

After analyzing around 500 news articles collected from NewsAPI with the question defined above, we conclude that the overall portrayal of Kamala Harris in the media is neutral, leaning towards negative, as over 38% of the articles have a negative connotation while only 15% are positive. However, this may be attributed to the fact that many articles focused on their disapproval of the campaign's budgeting or the alleged 1M USD paid to Oprah Winfrey. Since news outlets commit to maintain objectivity, it is unsurprising that 46.6% of the articles were neutral. However, there is a generally more negative

tone even in the supposedly neutral articles, which can be attributed to the loss of Kamala Harris in the election and thus, comes with a general feeling of disappointment from her supporters.

## Data

To answer the question of how Harris has been portrayed in the media since the election on November 5th, 2024, we collected a total of 535 articles using the API provided by newsapi.org, filtering with the keywords "Kamala" or "Harris" appearing in the title. To measure the general portrayal of Harris by the media, we decided to settle for the news outlets that had the biggest audience, since their headlines likely reach more of the population, even if they do not share the same political leanings. Hence, we sourced articles from the top 34 U.S. online news outlets according to *pressgazette.co.uk* (Press Gazette 2024) which we had access to through newsapi.org. For each relevant article, we obtained the title and the description. See Appendix (1) for the full list of chosen outlets.

We gathered all the articles available from the day following the election (November 6th) to the 21st of November. Empirically, we noticed certain trends that came with the articles being released closer or further away from the election night. After the first two days following the election, the articles predominantly consisted of live coverage of the election night and the results announcement, capturing the outcomes of different states, immediate reactions, and developments. In the subsequent week, expert analysis on why Harris lost to Trump and political commentary from other political figures began to emerge. These articles offered more in-depth perspectives and evaluations of Harris's role in the election and its outcomes. In the weeks of mid-November, much of the coverage shifted focus and started featuring topics less directly related to Harris herself, such as discussions around the broader election context, social media controversies, or celebrity opinions.

To refine the dataset, we excluded irrelevant articles, such as those mistakenly referring to other individuals with the same name, duplicate entries that listed articles with the

same title, and articles that appeared as “[Removed]”. This resulted in the deletion of 24 articles, leaving us with 511 articles, all of which were included in our analysis.

Another decision we made was to retain articles that were not explicitly focused on the election, as they still contributed valuable insights into how Kamala Harris was represented in the media during the post-election period. This comprehensive dataset includes a wide range of topics, such as real-time updates on election night, Harris’s concession speech, coverage of Trump’s victory celebrations, detailed expert analyses, public reactions, and broader discussions on post-election developments. By including all articles that mentioned Kamala Harris in the title, regardless of the primary subject, we aimed to capture a holistic view of her portrayal across different media narratives. This approach allowed us to observe patterns in how she was framed and contextualized in various political, social, and cultural discourses following the election.

As a result, our final data set is composed of 511 articles, each with the title of the news article, the description (the subtitle or the first sentence of the body), the category (the human-annotated inferred feature labelled according to the typology defined - see *Methods*) and the tone (the human-annotated inferred feature referencing the general positive/negative/neutral tone of the article - see *Methods*).

### **Minimizing biases:**

Before collecting our data, we also discussed the possibility of biases in our collection process. Since bias is unavoidable, we considered ways to minimize unwanted biases and work with the present ones. The following biases were taken into account in our article source selection: the political leaning of the news outlets and the period over which we gathered articles.

To respect our data science question, we had to include only English articles published in American news outlets. Then came the question of a political stance bias; many popular news outlets are known to lean left or right and thus tend to publish articles that benefit their agenda. However, as mentioned previously in *Data*, these popular outlets reach the majority of Americans not depending on their political stance and thus, together, create an umbrella of news that is representative of what influences the general public.

As for the dates, we decided to collect data from November 6th to 21st, which gives us a spread-out window of articles covering topics including election night, political criticism, and social media controversies. The 1-month lookback limit on NewsAPI did not affect our research, as we specifically wanted articles that were published after the election, which most likely impacted the tone of the authors.

## **Methods**

The methodologies we used in our experiment involve the following: development of the typology and tone definition, TF-IDF analysis, and tone analysis.

### **Developing a typology and defining the tones:**

For our research, we annotated two inferred features: *category/topic* and *tone (positive/negative/neutral)* of the articles (see *Data*). To develop a refined and comprehensive typology, we started with open coding rounds on 200 articles and reached a final typology after three rounds. Dividing the time period into three brackets, we took approximately 66 articles from each to obtain a comprehensive typology over our data collection period (the time brackets were November 6th-9th, November 10th-13th, and November 14th-21st; the final one being longer since outlets published more articles about Harris closer to November 6th and less as time went on). In the first round, we individually annotated a third of the 200 articles to brainstorm labels. This resulted in a plethora of different categories that we narrowed down to 11 labels. In the second round, we went through the 200 posts together and annotated them as a group, which led us to refine the definitions for the categories to reach a consensus. There were a lot of ambiguities raised and we renamed certain categories and made their definitions more comprehensive before finally settling on a final topology consisting of 7 categories which we used in our final round of coding on all collected articles.

In terms of the tone of the articles regarding Harris, since we did not need to go through the open coding process, we simply discussed and settled on definitions for classifying articles as either “positive”, “negative”, or “neutral”. For this process, we each took our sections to annotate and wrote down how the contents of the article based on the title and description are likely to impact readers' view of Harris. We concluded that the connotation attached to the overall tone and choice of words used to describe either Kamala Harris as a person or as a candidate was what mattered. Moreover, because we were annotating based on only article titles and descriptions, we decided to assume that articles talking about a specific person’s opinions had the same tone as the opinion itself unless otherwise implied. We decided that any article simply stating objective facts, such as: “Polls are starting to close across the U.S.”, would be neutral in tone. This also includes articles that cover mundane updates on her life post-election and those analyzing reasons why Harris lost, without taking a clear stance on whether they believe it was a good or bad thing.

### **TF-IDF analysis:**

We designed an algorithm to calculate the top 10 most relevant terms in each category by TF-IDF score, ignoring

stopwords. In cases where multiple words had the same 10th-highest score, we kept all such terms.

Using Matplotlib, we constructed bar charts for each category to visualize the TF-IDF scores for the top ten words in that label (see *Results*).

### Tone analysis:

Once we had annotated all the articles with topics, calculated the percentages of each category, calculated the TF-IDF scores, and constructed the visualizations, we moved on to the second annotated feature - the tone. As described above, once we decided on a definition for the tones, we annotated the articles and then grouped these results by category to obtain a tone breakdown by category. Similarly to the TF-IDF analysis, we used Matplotlib to create pie charts (Figures 8-15) representing the percentage of tones for each category (see *Results*). These visualizations helped us make the proportions and results more intuitive for the analysis.

## Results

Through the open coding process, we developed the following seven categories of our typology (see “Methods”):

1. **Concession Speech and Transition:** Articles about the concession speech or the transition of power after the confirmation of the election results. This category includes transitions of power and the transition in Harris’ career.
2. **General Public Responses and Reactions:** Articles mentioning any responses or reactions towards the election outcome coming from the general public. This includes non-politicians, supporters or non-supporters and the general public who are not politically involved.
3. **Post-Election Analysis:** Articles exploring why or how Harris lost the election, including analyses of campaign successes and failures, candidate critiques, demographic voting patterns, comparisons to past elections, and broader political trends.
4. **Election Night Coverage and Projections:** Articles covering the election night events (before and briefly after the results are announced) including state-by-state results roll-outs, predictions, and vote speculations and counts.
5. **Democrats’ Responses and Reactions:** Articles talking about general democratic party responses or reactions towards the election results. These come from internal personnel and group members and involve a person from the party sharing an opinion on behalf of the party as a whole.
6. **Political Commentary:** Articles featuring well-known politicians, political commentators, or

experts quoting their opinions on various aspects of the election.

7. **Gossip and Scandals:** Articles with controversial, click-bait titles featuring shock factors or large numbers, including celebrity and social media drama, post-election activities of political figures, and general election-related gossip not directly tied to the election itself.

Using the TF-IDF scores calculations, we obtained the following results for the top ten words in each of our seven categories:

**TF-IDF Scores of the top 10 words per Typology Category**

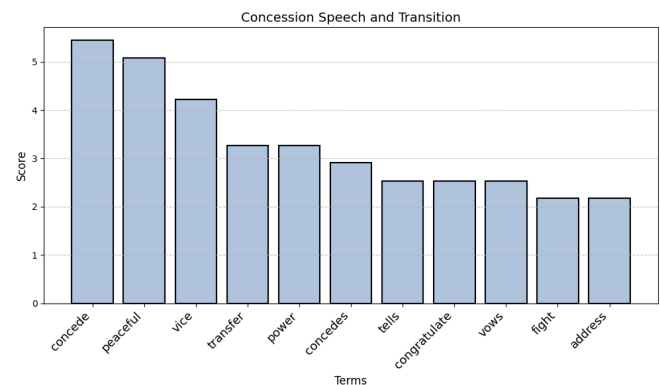


Figure 1: TF-IDF scores of Concession Speech and Transition

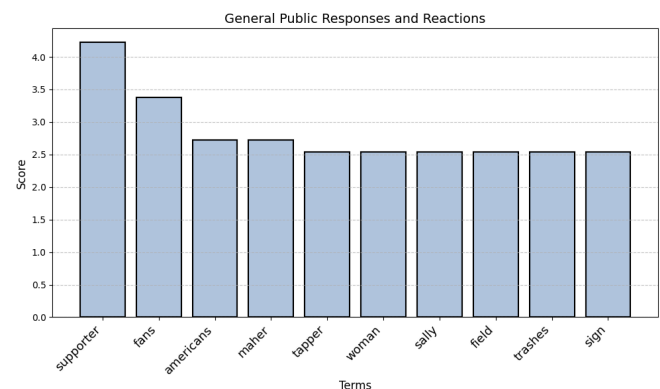


Figure 2: TF-IDF scores of General Public Responses and Reactions

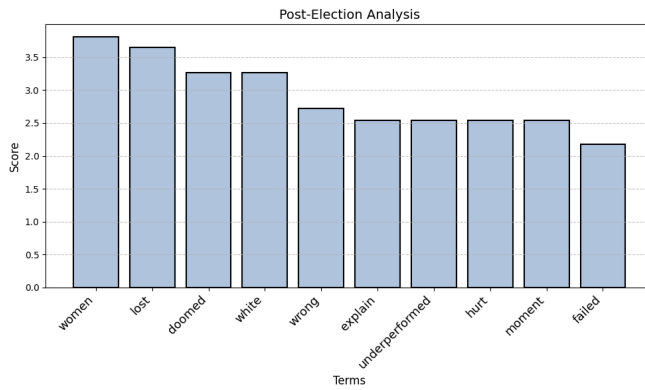


Figure 3: TF-IDF scores of Post-Election Analysis

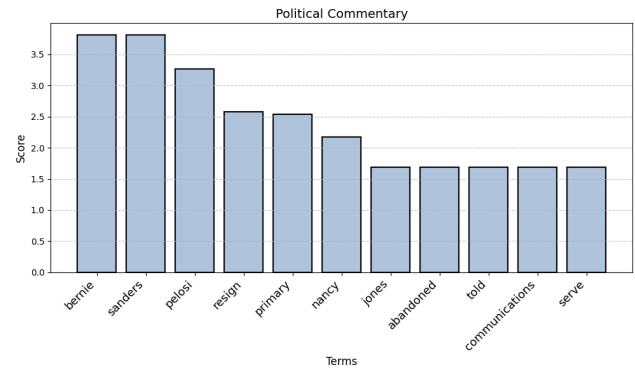


Figure 6: TF-IDF scores of Political Commentary

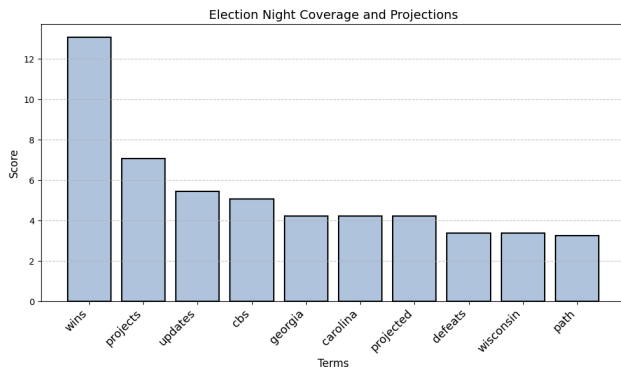


Figure 4: TF-IDF scores of Election Night Coverage and Projections

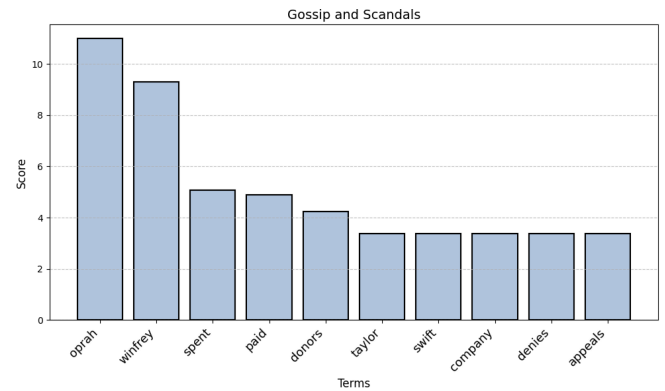


Figure 7: TF-IDF scores of Gossip and Scandals

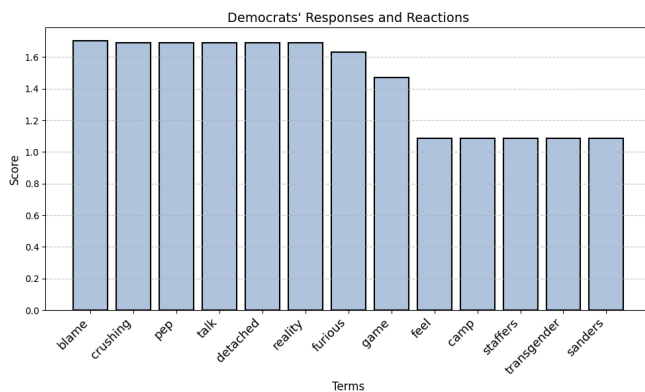


Figure 5: TF-IDF scores of Democrats' Responses and Reactions

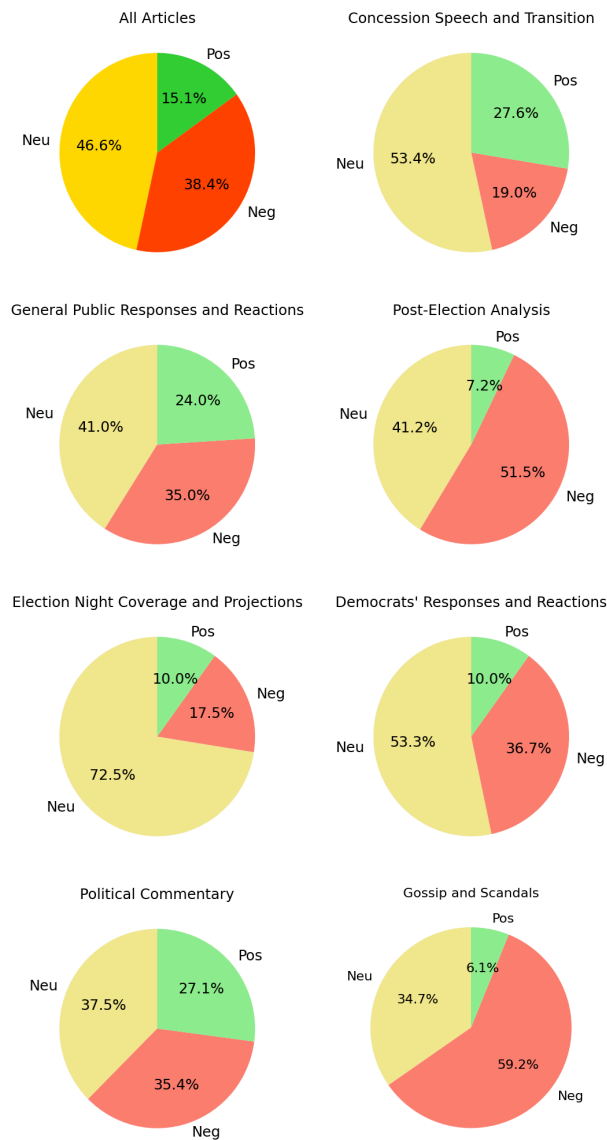
Observing the bar charts, we noticed that the distributions of the TF-IDF scores seem fairly uniform for the *General Public Responses and Reactions*, *Democrats' Responses and Reactions* and *Post-Election Analysis* but it is quite skewed when it comes to the *Concession Speech and Transition*, *Gossip and Scandals*, *Political Commentary* and *Election Night Coverage and Projections* (Figures 1-7).

Through observation, the top-scoring words do not necessarily have a positive or negative connotation. While some of them can easily be inferred as positive, such as “concede” and “peaceful” in *Concession Speech and Transition*, or negative, such as “blame” and “crushing” in *Democrats' Responses and Reactions*, those with neutral terms can still be categorized as negative given the context. For example, we find the words “bernie” and “sanders” in *Political Commentary*. However, the headlines involving Bernie Sanders mostly cover his disapproval of the way the Democratic Party ran their campaign.

Moreover, we also calculated the percentage of each tone (“positive”, “neutral”, or “negative”) in each of our seven categories (see *Methods* for the definitions) and for all the articles. The overall results show that 38.4% of articles have a negative tone, 46.6% are neutral and 15.1%

are positive (Figure 8). The following pie charts show the proportions overall and in each category:

### Tone Distribution Across Typology Categories



Figures 8-15: Charts showing the percentage of articles by tone overall and in each category.

## Discussion

Looking at our results, many of the top-scoring words in terms of TF-IDF scores matched what we expected given the context, such as “supporter” in the *General Public Responses and Reactions* category (Figure 2) or words such as “blame”, “crushing”, “furious” in the *Democrats’ Responses and Reactions* section (Figure 5), as the outcome of the election were surely disappointing for the

Democratic party and its supporters. This aligns with observations from our second annotation feature, tone, where many articles exhibited a negative tone, as reflected in the prevalence of these higher-scoring words. Some interesting results we found are discussed below.

The *Gossip and Scandals* category predictably held the highest proportion of negatively toned articles at 59.2% (Figure 15), compared to the six other categories with lower negatively toned articles, scoring around 20-30%. These articles focus on, as the name suggests, activities that involve Harris that are not always about the election. For example, some of the most prevalent articles were about the 2.6M USD spent on private jets in the final weeks of the election or mundane things about her first appearance with Biden since November 6th. This category also included celebrity drama she was part of or mentioned in or any exaggerations of the truth. In our case, quite a few of our articles from November 11th-15th were related in some way to accusations against Harris for spending one million dollars on celebrity Oprah Winfrey to endorse her campaign. This is reflected in our TF-IDF scores where “oprah”, “winfrey”, “spent”, and “paid” are the highest-scoring words in that category (Figure 7). Extrapolating from this finding, we do expect that if we collected data from a few days later, the main scandal topic would have moved on from Oprah Winfrey to whichever drama Harris became part of.

On the other hand, the category with the highest proportion of positively toned articles was *Concession Speech and Transition*. However, it is important to note that only 27.6% of these articles were positive, with the majority, 53.4%, being “neutral” (Figure 9). The other categories were predominantly neutral and negative, meaning that *Concession Speech and Transition* was the least negative of all categories. The TF-IDF analysis reveals “peaceful” as the second most prominent word (Figure 1). Kamala’s commitment to a peaceful transfer of power may stand out as a positive contrast to Trump’s refusal to ensure one in 2020 (Center for Presidential Transition 2022).

*Post-Election Analysis* is the second most negative category with 51.5% of negatively toned articles (Figure 11). It should not come as a surprise, since this category encompasses mostly expert or other qualified personnel analysis of why the Democratic campaign failed and Harris lost. The top-scoring words according to the TF-IDF are “women”, “lost” and “doomed” (Figure 3). This suggests that there was significant interest in the possibility of electing the first female president. It also reflects that there was heightened interest in how women in particular voted, considering this.

There was a predominantly neutral category: *Election Night Coverage and Projections* (Figure 12). For this section, the TF-IDF scores revealed typical words, such as “wins” and “projects” (Figure 4), which suit the general theme of the category, since they were mostly reports of

the current events as the election night unfolded into the early mornings of the 6th, without much room for left or right leaning biases.

Overall, the media coverage of Kamala Harris during this period primarily focused on the election, as it was the most recent and significant event she participated in that had a big influence on how the media portrayed her. However, as with any prominent public figure, the media coverage of Harris inevitably extended beyond just the election to encompass any “hot topics” or controversies she may have been involved in at the time. These topics, whether political or personal, often take the spotlight, adding to her public image.

In analyzing the general tone of the media coverage, it is clear that a substantial portion of articles (46.6%) adopted a “neutral” tone (Figure 8). This likely reflects the media’s strive for unbiased reporting, especially when covering high-profile figures like Harris. However, given that Harris did not win the election and lost the popular vote, it is unsurprising that a significant portion of the coverage (38.4%) was negative or critical of her campaign (Figure 8). This critical tone likely stems from her being a key figure in a losing campaign.

Even though many of the news outlets in our sources are generally left-leaning by observation, they do not seem to portray Harris in a more favourable light. On the contrary, the prevailing sentiment in the coverage leans more towards disappointment and blame, rather than appraisals or focusing on any of the positive aspects of her campaign. This highlights the interesting dynamic between prominent political figures and media bias independent of their political stance.

## Group Member Contributions

Jacqueline Wu: question definition, open coding, category definitions, category annotation, tone annotation, “data”, “results”, “discussion” and editing.

Erin Wyndham: question definition, data collection scripts, data collection, open coding, category definition, category annotation, tone annotation, TF-IDF, proportions of tones in categories, visualizations, “discussion” and editing.

Greta Ru-Mei Zu: question definition, data collection scripts, open coding, category definition, category annotation, tone annotation, “introduction”, “methods”, “results”, “discussion”, editing and formatting.

## Appendix

- (1) The following are the 34 American news outlets accessible via the newsapi.org API from the top 34 news outlets according to pressgazette.co.uk:
  - people.com

- yahoo.com
- nypost.com
- forbes.com
- cnbc.com
- cnn.com
- foxnews.com
- usatoday.com
- washingtonpost.com
- newsweek.com
- apnews.com
- cbsnews.com
- nbcnews.com
- businessinsider.com
- wsj.com
- huffpost.com
- substack.com
- reuters.com
- buzzfeed.com
- thehill.com
- politico.com
- breitbart.com
- variety.com
- sfgate.com
- dailydot.com
- zerohedge.com
- latimes.com
- rawstory.com
- nj.com
- thedailybeast.com
- msnbc.com
- today.com
- theatlantic.com
- the-sun.com

## References

Center for Presidential Transition. 2022-01-20. The 2020-21 Presidential Transition. <https://presidentialtransition.org/reports-publications/2020-21-lessons-learned/>. Accessed: 2024-12-02.

Press Gazette. PressGazette. 2024-10-10. Top 50 news websites in the US: All but four sites saw traffic fall in September. <https://pressgazette.co.uk/media-audience-and-business-data/media-metrics/most-popular-websites-news-us-monthly-3/>. Accessed: 2024-10-16.