Trump VS Everyone: An Analysis of Emotional Rhetoric in US Presidential Debates

Text Mining and Sentiment Analysis
Master in Data Science for Economics

Academic Year 2024/25

Stefano Chiesa
Università degli Studi di Milano

May 27, 2025

1 Introduction

The strategic use of emotional language in political discourse is a powerful tool for shaping public perception and influencing voter behaviour. This project explores how emotion is deployed rhetorically in US presidential debates. The objective is to classify speeches based on emotions and interpret the candidate's rhetorical goal. The full code is available at GitHub.

2 Research Question

How are specific emotions used by Trump and his opponents during presidential debates to frame the discourse and influence the audience? Is there a discernible pattern through the years in emotional usage across political affiliations and debate contexts, based on the status of the candidate?

3 Motivation and Critical Choices

The decision to focus on US presidential debates was motivated by personal interest and the broad range of topics these events cover, making it possible to observe how emotional expressions vary across different subjects. A key methodological choice was the adoption of the SamLowe RoBERTa-based model, a transformer-based classifier fine-tuned from roberta-base on the GoEmotions dataset, which is used to detect and categorise emotional tone. The model was selected due to its widespread adoption, as evidenced by its download frequency in the last month (over 500k downloads), suggesting a high level of community trust. In addition, the authors of this model were the only ones, among the large models available, to provide a detailed performance breakdown. Their evaluation shows strong performance on common emotions such as *gratitude*, *love*, and *amusement*, while highlighting challenges in detecting subtler or underrepresented emotions like *relief*, *disappointment*, and *realization*—a pattern that is crucial to consider when interpreting the model's outputs. In addition, the RoBERTa training dataset is over 10 times the size of the dataset used for BERT. The GoEmotions dataset includes 27 fine-grained emotion labels plus a neutral class. Although this level of granularity is potentially informative, I opted to aggregate the emotions into broader categories to facilitate the interpretability of the results. A similar decision process was followed for choice of the topic classification model (dstefa), a RobBERTa-base model fine-tuned on the NYT News dataset, which contains 256,000 news titles from articles published from 2000 to the present (Table 1). The performances are much better than the next model, probably because the dataset is more balanced and it is easier to assess topics than emotions. The models were run on an NVIDIA 1070 Ti, using CUDA, since it is much faster than CPU execution.

4 Methodological Framework

4.1 Data Collection and Cleaning

The transcripts were sourced from the *American Presidency Project* website, specifically from the debates section covering the years 2016 to 2024. This includes debates involving Donald Trump, spanning his roles as a private citizen, sitting president, and candidate seeking re-election. Joe Biden also participated in different years, although to a more limited extent. This temporal and contextual diversity is crucial, as it enables an analysis of behavioural differences across various campaign phases and political roles, and allows for the identification of potential rhetorical patterns over time.

4.2 Transcript Processing

Each debate transcript was named following the format `SPEAKER1_SPEAKER2_LOCATION_YEAR.txt`, from which metadata were extracted automatically. A custom script parsed the dialogues, identified speakers, and classified unknown ones as "Moderator". Speeches were cleaned by removing annotations such as `[applause]` and `[laughter]`, which do not contribute meaningful semantic content for classification tasks.

4.3 Segmentation

To facilitate emotion and topic analysis, speech texts were segmented into chunks of up to 30 words, respecting sentence boundaries where possible. When an answer exceeded this limit, it was split into smaller parts without disrupting syntactic structure. This step was necessary because overly long segments often conveyed multiple emotions, reducing classification accuracy and interpretability.

4.4 Emotion Classification

Emotion detection was carried out using the `SamLowe/roberta-base-go_emotions` model, which assigns a single emotion label to each speech segment. These labels were subsequently grouped into broader categories to simplify interpretation and reduce noise (Table 2). This approach enables the analysis of the emotional tone adopted by each speaker across time and debate contexts.

4.5 Topic Classification

The topic classification step using `dstefa/roberta-base_topic_classification_nyt_news` was essential to better understand which topics were driving specific emotions in the speeches, allowing for an analysis of the relationship between content and emotional tone.

4.6 Explainability Methods

To interpret and validate the predictions of the emotion classification model, two complementary explainability techniques were employed: attention analysis and SHapley Additive exPlanations (SHAP). The first one is based on a qualitative analysis of the attention matrix weights, which should be higher for important words in determining the context of a sentence. The problem is that sometimes irrelevant tokens are highlighted (Chefer et al., 2021); in addition, higher weights do not imply a causal relationship between the input token and the output label of the model. On the other hand, SHAP is a unified additive feature attribution framework proposed by Lundberg and Lee (2017), which interprets complex models by assigning Shapley values to input features. First, we examined the attention weights from the `roberta-base-go_emotions` model to gain insight into which tokens the model focused on during prediction. Although attention is not a direct measure of feature importance, it provided a preliminary qualitative understanding of how emotional cues were distributed across each speech segment. For a more robust and model-agnostic explanation, we computed SHAP values on a subset of segments. SHAP quantifies the contribution of each token to the predicted emotion label by estimating local feature attributions. This allowed us to identify specific words or phrases that had the strongest influence on the model's output.

5 Result Analysis and Interpretation

The final stage involved the visualisation and interpretation of results using explanatory statistics. This includes tracking the distribution of emotions and topics over time, as well as identifying potential shifts in rhetorical strategy between different roles (Eg. incumbent vs challenger, Trump VS Democrats...).

5.1 Emotional trends

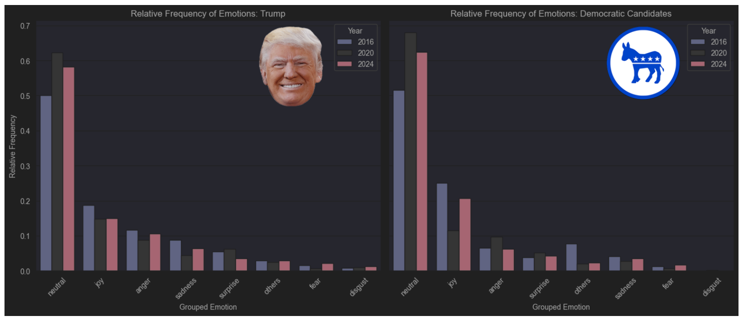

Figure 1: Overall distribution of emotional expressions in presidential debates, segmented by speaker and year.

On average, candidates who are not in office tend to express more anger and less joy. When we compare Trump's rhetoric in 2020 to the other years he was out of office, this pattern becomes evident. A similar trend can be observed among Democratic candidates as well. Trump typically exhibits higher levels of anger compared to his Democratic counterparts, alongside a decrease in expressions of joy and an increase in sadness (Figure 1).

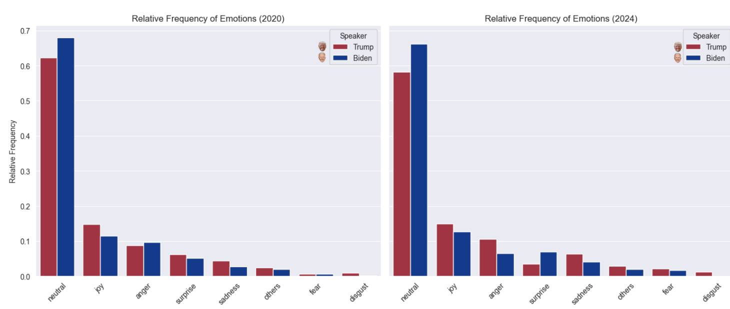

Figure 2: Comparison of emotional tone between Trump and Biden during debates.

The observation can be made that, in general, President Joe Biden appears to place limited emphasis on the concept of joy within his public discourse and policy initiatives (Figure 2). The context of his presidency, marked by significant challenges such as the COVID-19 pandemic and economic recovery, may further contribute to this impression.

5.2 Which are the topics that are more related to anger?

The objective is to identify the topics that elicit a more aggressive response from a given candidate.

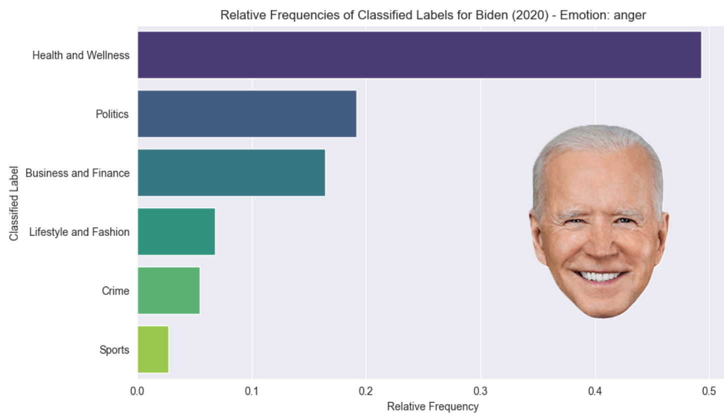

Figure 3: Topics most associated with anger in Biden's 2020 debate segments.

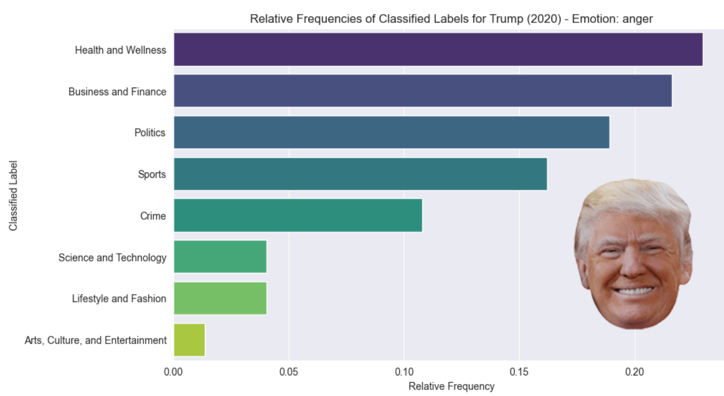

Figure 4: Topics most associated with anger in Trump's 2020 debate segments.

It is evident that in 2020, health emerged as the most critical topic of discussion, largely due to the ongoing pandemic (Figures 3 and 4).

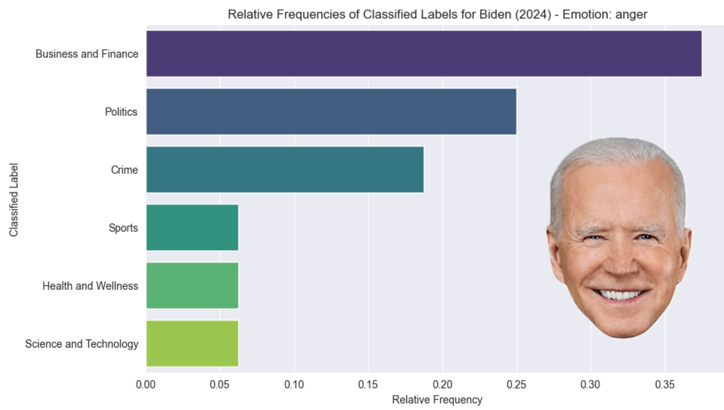

Figure 5: Topics most associated with anger in Biden's 2024 debate segments.

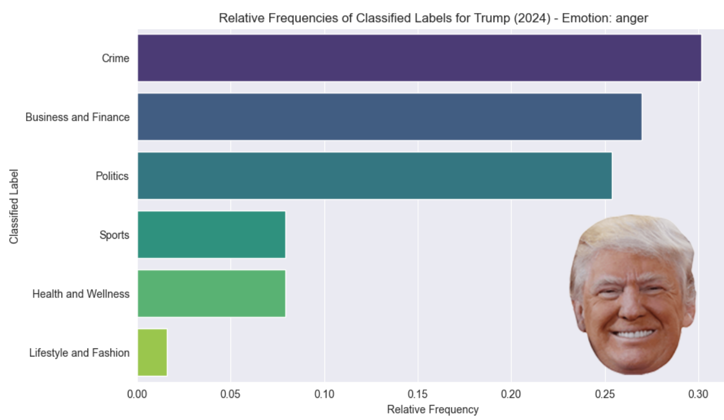

Figure 6: Topics most associated with anger in Trump's 2024 debate segments.

In 2024, crime emerged as the most discussed issue (Figures 5 and 6). The ongoing conflict in Ukraine was classified by the model as ”crime”; since it was a central topic in 2024 debates, this classification appears appropriate. Following crime, the topic ”business and finance” was the most predominant in Trump's rhetoric. Inflation hit hard globally during Biden's tenure, prompting Trump to focus on this issue, as it impacts a broad audience and can elicit a strong reaction. On the other hand, President Biden tended to avoid heated discussions about economic issues, likely due to his administration being in power during the inflation rise.

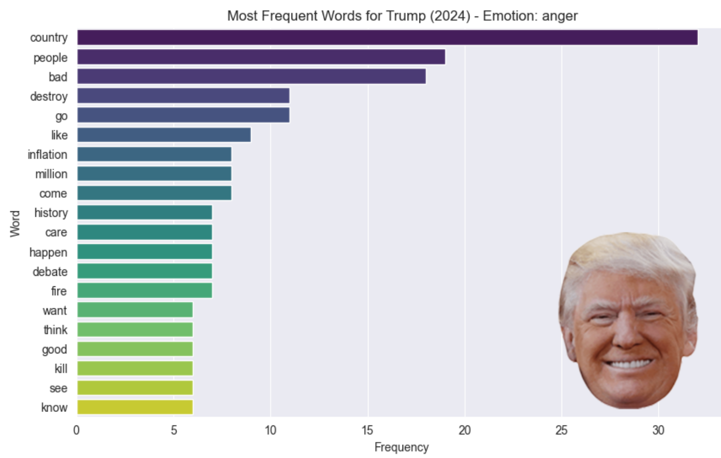

Figure 7: Words most associated with anger in Trump's 2024 debate segments

When examining the most commonly used word by Trump when he conveys anger, the first meaningful word is "inflation" (Figure 7). In contrast, drawing conclusions for Biden is more challenging, as he uses "anger" sentences less frequently (Figure 8).

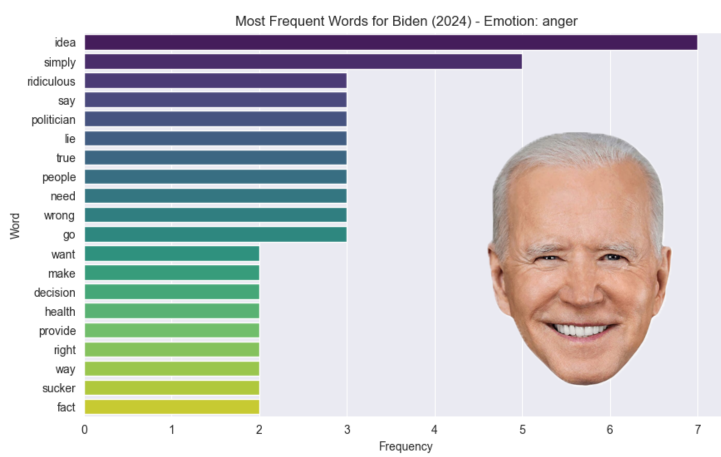

Figure 8: Words most associated with anger in Biden's 2024 debate segments.

5.3 Explain unexpected pattern

When examining the predominant emotion associated with Donald Trump in the context of the 2024 election, when he uses the word "economy," the emotion identified is "joy", followed by "neutral", and then by "sadness". This finding presents a contrast to earlier findings, which suggested that individuals outside

of the office tend to adopt a more aggressive stance on significant issues. An attempt was made to explain this discrepancy by reading the sentences (Table 3 and Table 4), analysing the attention matrix and using the SHAP explainer method.
A human can understand that Trump refers to his economy when being joyful, and to Biden's economy when being sad. Consider, for example, the sentence `"The greatest, before COVID came in, the greatest economy in history, lowest employ--unemployment numbers, everything was good. Everything was going. And by the way, there was unity going to happen."`

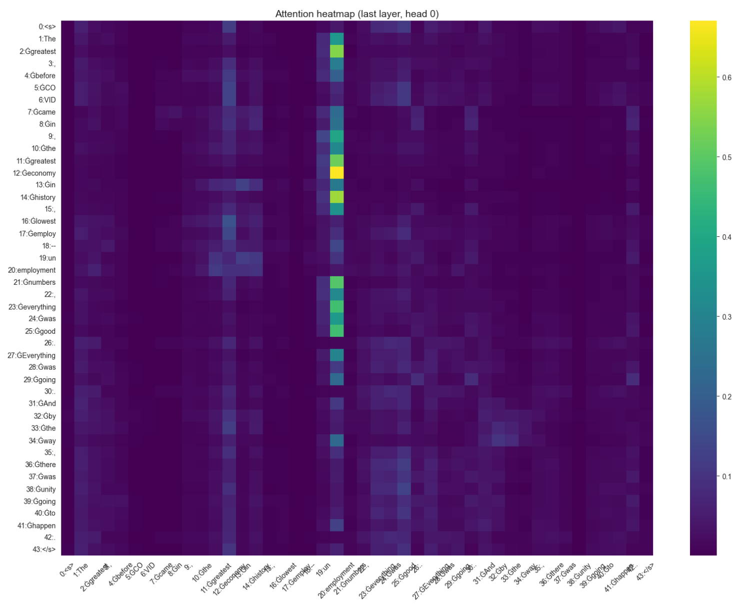

Figure 9: Attention matrix highlighting the focus around economic terms in a Trump segment classified as admiration.

We observe that the attention mechanism focuses on the context around the word "employment" (Figure 9). However, this attention pattern alone does not fully explain why the model predicts the emotion "admiration" (mapped to "joy" in our categorisation). It is important to note that the attention weights do not represent a causal influence on the model's final output; rather, they provide insights into how the model contextualises the input tokens, which might affect the output of the model.
SHAP reveals that words such as *"good"* and *"greatest"* (appearing twice) have a significant positive impact on the model's classification (Figure 10). This can be interpreted as a causal relationship.

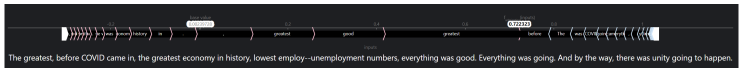

Figure 10: SHAP values indicating the contribution of each word to the classification of “admiration” in Trump’s segment.

The sentence `"Well, look, the greatest economy in the world, he’s the only one who thinks that, I think."` from Biden is clearly an ironic statement. He is referring to the fact that Trump considers his the economy the best, and he does not agree. According to the emotion classification model, the emotion of the sentence is ”Joy”, which is a mistake.

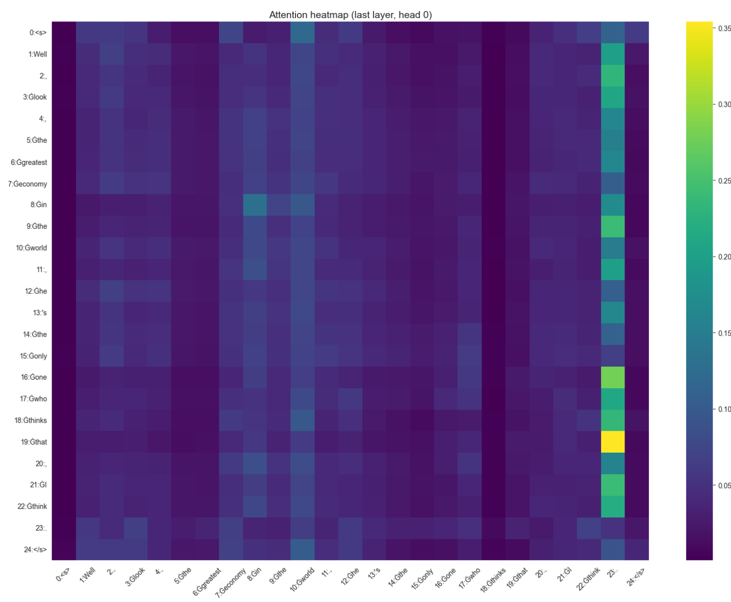

Figure 11: Attention matrix for Biden’s ironic statement about Trump’s view of the economy.

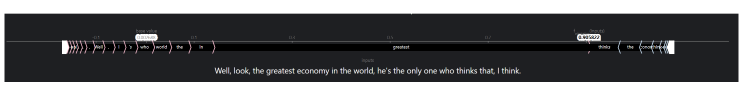

Figure 12: SHAP values indicating the contribution of each word to the classification of “admiration” in Biden’s segment

Looking at the attention matrix, we cannot find any meaningful insights. In addition, we notice that the token ”.” is highlighted, even though it should be way less relevant than other words in terms of the meaning of the sentence, an eventuality I mentioned at the beginning (Figure 11). The SHAP reveals that the words ”greatest” push for the sentiment ” admiration” (Figure 12), which is then mapped to ”Joy”, similarly to what happened before. It looks like some keywords, probably because of an imbalanced fine-tuning dataset, have a huge impact on the classification decision. They overcome most of the contextual information that might tell the model the sentence is ironic or at least not joyful. This happens similarly with optimism (joy), with the pattern ”see what happens” (Figure 13).

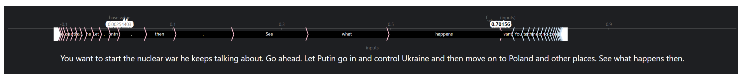

Figure 13: SHAP values indicating the contribution of each word to the classification of “optimism” in Biden’s segment.

6 Limitations and Concluding Remarks

Several limitations must be addressed to contextualise the findings. First, the `roberta-base-go_emotions` model demonstrates stronger performance for common emotional categories such as joy and gratitude, but performs poorly on less frequent or more nuanced emotions like relief or realisation. This performance imbalance may distort the emotional landscape presented in the analysis, especially if key rhetorical strategies rely on underrepresented emotional cues. Second, the topic model was fine-tuned on news data (NYT corpus) rather than political speech, and domain mismatches—such as framing international conflict as “crime”—highlight the risk of misclassification that can affect downstream interpretations. The temporal and speaker coverage of the dataset also presents limitations: debates span only from 2016 to 2024 and include uneven speaker participation, making trend comparisons more difficult. Then, the choice of 30-word chunks was a practical compromise, but different segmentation thresholds may yield different emotional interpretations. In addition, the rhetorical patterns weren’t highlighted with a structural method like in Şeref et al. (2023), which relies on manually annotated rhetorical moves and a sequence-alignment algorithm to generalise them across large corpora. This approach, while powerful, requires a predefined set of rhetorical categories, expert-labelled training data, and significant manual effort, which was not feasible within the scope and timeframe of this project but it is for sure an idea for future work. Then, it is important to recognise the potential for unintended author bias. Despite efforts to maintain analytical neutrality, prior expectations—such as anticipating aggressive behaviour from specific candidates—may have subtly influenced interpretative choices or the framing of conclusions.

7 AI Usage Disclaimer

Parts of this project were developed with the assistance of OpenAI's ChatGPT (GPT-4 and Claude 4). The AI was utilized to help generate ideas, support coding development, and draft the Overleaf template. Additionally, Grammarly AI was employed to enhance the quality of the language in the text, as I am not a native English speaker. Every piece of content produced with AI assistance has been thoroughly reviewed, edited, and validated by me. I assume full responsibility for the final content, ensuring its accuracy, relevance, and adherence to academic integrity.

References

Chefer, H., Gur, S., & Wolf, L. (2021). Transformer interpretability beyond attention visualization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://github.com/hila-chefer/Transformer-Explainability

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*. https://github.com/slundberg/shap

Şeref, M. M. H., Şeref, O., Abrahams, A. S., Hill, S. B., & Warnick, Q. (2023). Rhetoric mining: A new text-analytics approach for quantifying persuasion. *INFORMS Journal on Data Science*, *2*(1), 24–44. https://doi.org/10.1287/ijds.2022.0024

A Appendix

Table 1: Performance of dstefa model for topic classification

Topic	Precision	Recall	F1-score	Support
Sports	0.97	0.98	0.97	6400
Arts, Culture, and Entertainment	0.94	0.95	0.94	6400
Business and Finance	0.85	0.84	0.84	6400
Health and Wellness	0.90	0.93	0.91	6400
Lifestyle and Fashion	0.95	0.95	0.95	6400
Science and Technology	0.89	0.83	0.86	6400
Politics	0.93	0.88	0.90	6400
Crime	0.85	0.93	0.89	6400

Table 2: Performances of SamLowe model for emotion classification

Grouped Emotion	Original Labels
Joy	amusement, approval, excitement, gratitude, joy, love, optimism, pride, relief
Anger	anger, annoyance, disapproval
Sadness	disappointment, grief, remorse
Fear	nervousness
Disgust	disgust
Surprise	surprise, realization, confusion, curiosity
Others	caring, desire, embarrassment, etc.
Neutral	neutral

Table 3: Sentences containing the word *economy* in speeches by Trump in 2024, labelled with emotion *joy*

#	**Sentence**
1	We had the greatest economy in the history of our country. We have never done so well. Every – everybody was amazed by it. Other countries were copying us. We got hit with COVID.
2	We got a lot of credit for the economy, a lot of credit for the military, and no wars and so many other things. Everything was rocking good.
3	I made great trade deals with the European nations, because if you add them up, they're about the same size economically. Their economy is about the same size as the United States.
4	We have to get 'em out fast. I created one of the greatest economies in the history of our country. I'll do it again and even better.
5	But it makes no difference. I have nothing to do–everybody knows I'm an open book. Everybody knows what I'm going to do. Cut taxes very substantially. And create a great economy like I did before.
6	We had the greatest economy. We got hit with a pandemic. And the pandemic was–not since 1917 where 100 million people died has there been anything like it.
7	We did a phenomenal job with the pandemic. We handed them over a country where the economy and where the stock market was higher than it was before the pandemic came in.

Table 4: Sentences containing the word *economy* in speeches by Trump in 2024, labelled with emotion *sadness*

#	**Sentence**
1	They've destroyed the economy and all you have to do is look at a poll. The polls say 80 and 85 and even 90% that the Trump economy was great that their economy was terrible.
2	And just look at what they're doing to our country. They're criminals. Many of these people coming in are criminals. And that's bad for our economy too.
3	I built one of the greatest economies in the history of the world and I'm gonna build it again. It's going to be bigger, better and stronger. But they're destroying our economy.