Has US Diplomatic Discourse Changed? An Analysis of Trump vs. Biden OSCE Statements

Alin Ierima

2025-04-17

1. Introduction

This paper investigates whether US diplomatic discourse has changed between the Trump and Biden administrations by comparing OSCE statements. Using quantitative text analysis, we extract features via bag of words, TFIDF, and bigram representations, then employ supervised classification (Naive Bayes) to test if the speeches are distinguishable.

In a final paper for another class (link), I have explored the implications of the US administration change to its foreign policy, by engaging with the diplomatic statements from the weekly Permanent Council meetings. By using discourse analysis methodologies, the paper revealed a significant rhetorical shift from a moralised, value-based framing of the war to a procedural, depersonalised, and strategically neutral tone following Trump's return to office.

In the present paper, I extend this analysis by applying quantitative text analysis methodologies to corroborate the qualitative findings and to increase the sample size (from 6 texts to 24). I expect, and hypothesise, that texts released after the cutoff point of January 20 will exhibit a distinguishable difference in word usage compared to earlier statements.

Given the final paper requirements focus on the analysis part, the rest of the introduction will discuss methodology. In order to compare the statements between the two periods in an "apples to apples" way, I decided to only pick the US statements at the OSCE Permanent Council under the agenda "The Russian Federation's ongoing aggression against Ukraine", which became a recurrent feature of the meetings after the full-scale invasion started. Since only twelve such statements have been released under the Trump administration, the same number has been picked for the Biden statements so as to have a balanced dataset from the very beginning. For Biden, the statements range from 19th September to 19th December 2024, and for Trump, from 23rd January to 10th April 2025.

There have been a couple of challenges to overcome in terms of data availability. For example, statements are no longer made public on the US Mission's website after the Trump inauguration. Additionally, the automatic scraping of Biden statements from the website has been unsuccessful due to the anti-bot measures of the website. Therefore, all statements have been manually collected from the OSCE internal system.

The analysis will proceed as follows. First, a descriptive analysis will be made with the use of three text representation techniques. Bag of words, or word frequency, will be used to provide a general overview of the most commonly used words across the documents; a visual representation will also facilitate this part. Frequency-inverse document frequency, or TFIDF, will take a similar approach to bag of words, yet focus on words that are distinctive across the two eras (Biden and Trump), weighting them as more important the less common they are. Lastly, a bigram representation will focus on collocations, or sequences of two consecutive words, which should enhance the analysis by offering an insight into contextual and rhetorical

patterns that are recurrent in the statements. Taken together, these three feature representations provide a good image on whether there is a significant difference between the two eras, as hypothesised.

The second part of the analysis will take these feature representations and use supervised classifications (Naive Bayes) to train a model on half of the sample to distinguish between the texts from the two administrations on the other half. The objective here is to see whether these models can distinguish between the two types of texts, and if so, which feature representation is the most effective. For robustness reasons, a 10-fold cross-validation will be undertaken, which evaluates the overall classification performance with metrics such as accuracy, precision, recall and F1 score.

2. Data Cleaning, preprocessing

In the data cleaning phase, raw text extracted from the PDF documents was prepared for the preprocessing by removing elements that do not bring any value to the analysis such as boilerplate sections, excessive whitespace and possessive markers ('s). After cleaning, the data was transformed into a corpus to prepare for the subsequent transformations.

Subsequently, in the preprocessing step, the corpus was tokenised to convert the text into individual words and to remove punctuation. Finally, word stemming was applied to reduce each word to its base form.
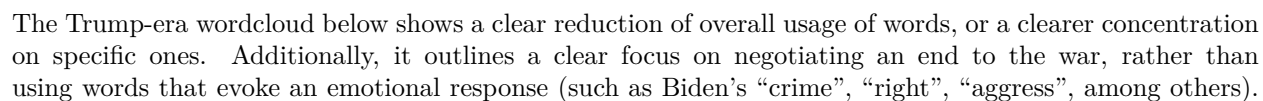
3. Descriptive analysis

To begin, I created a Document-Feature Matrix (DFM) from the tokenised corpus, which represents each document as a row and each unique word as a column. Then, the resulted matrix is used to identify the most occuring words in the dataset.

Table 1: Top 10 word frequency

Overall_words	Overall_freq	Biden_words	Biden_freq	Trump_words	Trump_freq
russia	155	ukrain	108	russia	52
ukrain	151	russia	103	ukrain	43
war	71	war	33	war	38
state	48	peopl	30	peac	27
secur	41	secur	28	presid	25
attack	40	support	28	end	24
unit	39	ukrainian	26	state	24
peac	38	state	24	must	23
peopl	38	attack	20	attack	20
support	35	unit	19	unit	20

Next, I extracted the top 10 overall frequent words from the entire corpus and also disaggregated the frequencies by eras. Table 1 shows that terms such as "russia", "ukrain" and "war" are popular across the eras, yet the Trump statements tend to have lower frequencies despite similar word counts - confirming one of the conclusions from the previous study that mentioning the two countries tends to be avoided overall. A visual representation of the difference between the two eras was created through word cloud visualisations (first is Biden-era, followed by Trump-era documents).

The Trump-era wordcloud below shows a clear reduction of overall usage of words, or a clearer concentration on specific ones. Additionally, it outlines a clear focus on negotiating an end to the war, rather than using words that evoke an emotional response (such as Biden's "crime", "right", "aggress", among others).

Next, I look at TFIDF, which weights words based on their inverse frequency. As expected, words such as "dprk", "justic", "iran", "human" and "crime" are more proeminent with Biden, as the statements tend to reflect a critical attitude towards the Russian actions and their collaborating allies. In contrast, Trump's statements feature "negoti", "ceasefir", "must", "presid", "achieve", which all strongly suggest a rhetorical focus that emphasises outcomes, leadership and a negotiation-oriented stance.

Table 2: Table 2: Top 10 TFIDF statements features

Biden_words	Biden_weighted_freq	Trump_words	Trump_weighted_freq
dprk	16.255620	ceasefir	8.559664
justic	9.712631	negoti	8.026698
iran	8.633450	must	7.792827
human	8.428840	presid	6.656697
soldier	8.174895	toward	6.421358
crime	7.781512	achiev	5.449930
food	7.554269	stop	5.418540
council	7.224720	minist	5.395906
kremlin	7.156819	end	4.898880
damag	7.003361	way	4.816480

In order to capture common pairings of words that may convey some deeper contextual meanings, I generated bigram tokens from the preprocessed text, which I then used to create a DFM that is dissagregated by administration. As shown in table 3, collocations such as "final act" - referring to the Helsinki Final Act, and "human right", outline the US' moralised approach during the Biden administration. During Trump however, the constructions in the statements primarily drew their legitimacy from peace as an ultimate

objective ("war must", "end war"), under the helm of Donald Trump ("presid trump").

Table 3: Table 3: Top 10 bigram features

Biden_words	Biden_freq	Trump_words	Trump_freq
unit_state	15	unit_state	20
final_act	10	presid_trump	13
osc_region	9	must_end	8
helsinki_final	9	russia_ukrain	7
russia_war	9	war_must	7
madam_chair	9	end_war	6
peopl_ukrain	9	final_act	6
support_ukrain	8	russia_war	6
russia_forc	7	death_destruct	6
human_right	6	presid_putin	5

The wordclouds further outline the stylistic and thematic differences between Biden-era and Trump-era diplomatic discourse. Most strikingly, collocation "presid trump" becomes central, despite the fact the diplomatic statement genre is usually more detached and avoids naming specific actors.

presid_trump

war_ukrain
madam_chair
final_act
war_must   must_stop
end_war   drone_missil

unit_state
must_end   presid_putin
russia_ukrain
russia_war
death_destruct
ukrain_russia

4. Supervised machine learning

Having explored the raw and weighted token distributions, now I turn to a formal test of the core question: can a simple classifier tell Trump-era from Biden-era statements based on word usage only? To do this, we use our three representations already computed (bag of words, TFIDF, bigrams) earlier. With only 24 total documents, I maximise the training data by splitting the corpus evenly, in 50% for training and 50% for testing. Additionally, I preserve the Trump/Biden ratio in each set to still be 50/50, so that the model will train on an equal number of era-statements.

I then fit a Naive bayes multinomial model on the training half. For the confusion matrix, I treat Trump as the "positive" class, because my focus is on detecting the new administration's rhetorical shift. I do the exact same thing for all three representations.

On the hold-out set,I continue by extracting four key metrics which are used to evaluate the models (accuracy, precision, recall, F1).As can be seen below, bag of words and TFIDF performed identically across all metrics, while bigrams outperformed both by achieving the highest accuracy (0.9) and perfect (1) precision. This is expected, as collocations such as "presid trump" and "cease fire" are exclusively used in the Trump-era documents.

Table 4: Evaluation Metrics by Feature Representation

Accuracy	Precision	Recall	F1
0.833	0.8	0.8	0.800
0.833	0.8	0.8	0.800
0.917	1.0	0.8	0.889

Despite the very similar performance metrics observed across the three models (BoW, TFIDF, and bigrams), these results do not indicate that the representations were redundant. The convergence in scores instead points to the stability of class distinctions in the dataset: a number of discriminative tokens (such as "Trump" "ceasefire" or "human", for example) appear to be strong enough to drive model predictions regardless of representation. To confirm that the representations were not identical and that the models were trained on different feature spaces I checked the actual features that were used:

```
identical(featnames(dfm_bow), featnames(dfm_tfidf_all))
```

```
## [1] FALSE
```

```
identical(featnames(dfm_bow), featnames(dfm_bi_all))
```

```
## [1] FALSE
```

5. Cross validation

To check the stability of the Naive Bayes classifiers, I performed a 10-fold cross-validation on each feature representation, using Laplace smoothing and stratified folds. In each fold, 90% of the statements serve as training data and 10% as test data. Then, the average across all ten folds is calculated.

```
## Average 10-fold Cross-Validation Accuracy for bag of words: 0.692
```

As can be seen, the bag of words achieved a 69% accuracy rate - still better than chance, but noticeably lower than the 83% that was present on the 50/50 hold-out split. This drop reflects the small size of the corpus and, therefore, its high variance. More data would have been necessary for improving the performance. TFIDF, as can be seen below, maintains an identical score.

```
## Average 10-Fold CV Accuracy for TFIDF: 0.692
```

Bigrams confirm their highest achieved performance, as already observed in the Naive Bayes hold-out results. With an average cross-validation accuracy of 75.8%, the bigram model outperforms both the bag of words and TFIDF representations. This might be due to collocational patterns that only emerge in certain eras - such as "presid trump" during the Trump era or "dprk soldier" in the Biden era.

```
## Average 10-Fold CV Accuracy for Bigram: 0.758
```

6. Instead of conclusion: discussion

The descriptive analysis confirms clear lexical and stylistic differences between the Biden and Trump administrations. As shown in the word frequency tables and wordclouds, Biden-era statements rely more heavily on moralised and rights-based vocabulary, such as "crime", "justic", "human", and "iran". These reflect a value-driven framing of international relations, particularly in response to the war in Ukraine. In contrast, Trump-era texts prioritise terms like "ceasefir", "must", "negoti", and "achieve", suggesting a more transactional, goal-oriented approach. This contrast is further reinforced by the bigram analysis, where collocations such as "presid trump" and "end war" dominate the Trump discourse, while phrases like "final act", "human right", and "support ukrain" are more prominent in the Biden texts.

The second part of the analysis showed that, across all three feature representations, the Naive Bayes classifier was able to distinguish between the two eras with relatively high accuracy. Bigrams outperformed both the bag of words and TF IDF models, reaching 90% accuracy compared to 83% for the others. A K-Fold validation, however, decreased that to a 75% accuracy, compared to 69% for the others. The fact that the two unigram-based models performed identically, despite using different weighting schemes and vocabularies, suggests that a small set of highly discriminative words, such as related to values, leadership and military actions, play an important role in separating the two periods, essentially functioning as shortcuts that consistently signal the administration behind each statement. This was expected.

There are, however, important limitations to the analysis. The dataset remains small, consisting of only 24 documents, which limits the robustness of the supervised classification results. While cross-validation helped mitigate this by testing model performance across multiple splits, results are bound to vary with the addition of new statements. Despite that, the existing analysis lends further credibility to the conclusions of my previous qualitative paper on this topic.