

VOLODYMYR ZELENSKY - THE VOICE OF THE UKRAINE

A PRESIDENT'S PLEA FOR NATIONAL RESILIENCE AND GLOBAL SOLIDARITY

INTRODUCTION

Since February 24th, 2022, Ukraine has been invaded by the Russian Federation. President Volodymyr Zelenskyy has revolutionized communication by using social media to engage with the Ukrainian people and the global community. Amidst the ongoing conflict, Zelenskyy has emerged as a prominent leader, representing the resilience of the Ukrainian people. Through strategic communication, he conveys the Ukrainian perspective worldwide, seeking stronger international support, legitimacy and recognition.

RESEARCH QUESTION

How has the thematic focus of President Zelenskyy's speeches to the Ukrainian people and world evolved from the initial stages of the war to the present?

DATA SET

- 24 recent speeches to Ukrainian / world (06/06 – 06/24 2023)
- The 24 first speeches to Ukrainians / world after war broke out (02/24 – 03/08 2022)
- Web scraped with Python from the official website of the President of the Ukraine: www.president.gov.ua/en

METHOD

1. **Preprocessing**
2. **Topic Model**
 - Latent Dirichlet Allocation (LDA)
 - Evaluation: grid search → coherence score
 - Separately trained models for both corpora
3. **Topic model comparison**
 - Jaccard distance

RESULTS

Corpus 1: The 24 first speeches to Ukrainians / world after war broke out

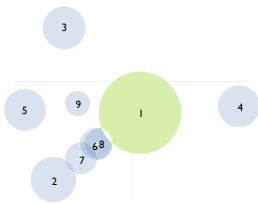


Corpus 2: 24 recent speeches to Ukrainians / world

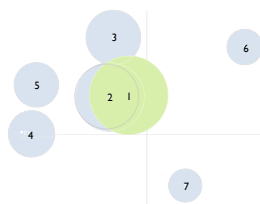


TOPIC MODELING

Intertopic Distance Map Corpus 1



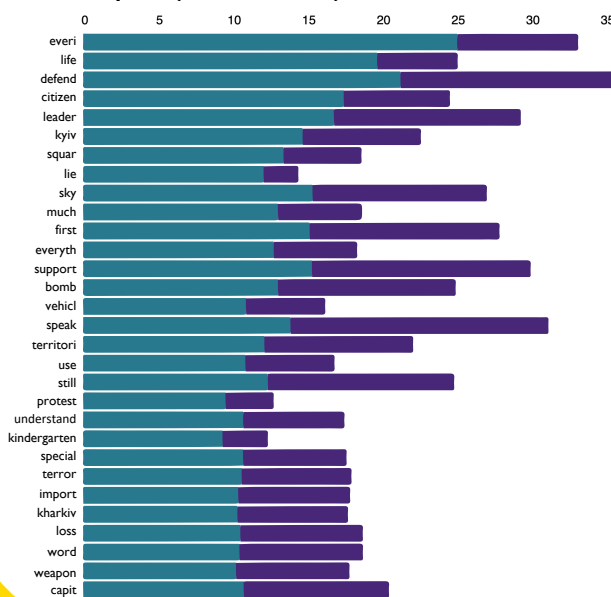
Intertopic Distance Map Corpus 2



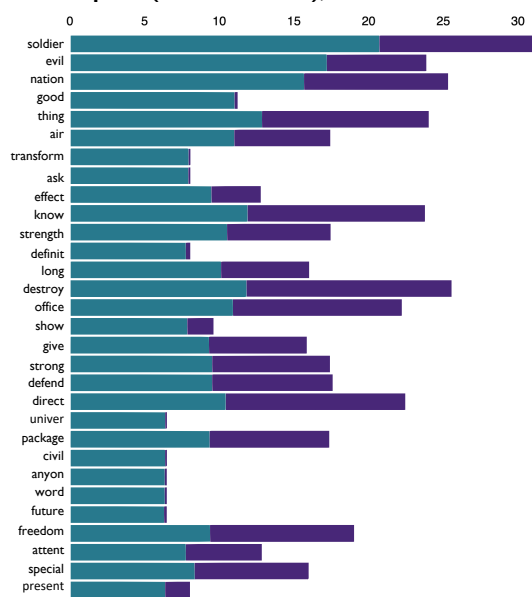
Jaccard Distance Heat Map

The Jaccard distance is calculated by finding the proportion of objects that are unique to each set compared to the total number of objects in both sets. A lower Jaccard distance indicates a higher similarity between topics, while a higher Jaccard distance suggests greater dissimilarity.

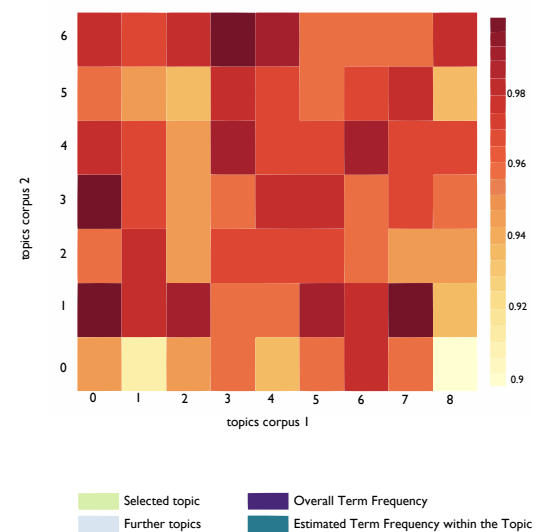
Top-30 Most Relevant Terms for Topic 1 in corpus 1 (38.9% of tokens), $\lambda = 0.61$



Top-30 Most Relevant Terms for Topic 1 in corpus 2 (30.3% of tokens), $\lambda = 0.61$



Topic difference between the Topic Models of Corpora 1 and 2



MODEL EVALUATION

- Hyperparameter tuning with grid search: parameters with the maximum coherence score were chosen → optimal number of topics is $k = 9$ for Corpus 1 and $k = 7$ for Corpus 2
- With hyperparameter tuning the coherence score went up from 0.3166 to 0.3541 for Corpus 1 and from 0.3493 to 0.3997 for Corpus 2

Topics	Alpha	Beta	Coherence
7	Asym metr.	0.01	0.4066
8	Symm etr.	0.91	0.3998
7	0.91	0.61	0.3969

Top three grid search results with the highest coherence score for Corpus 1

DISCUSSION

Topic 1 is the most prevalent in both corpora, making it the chosen representative of the topics discussed within each corpus. Nonetheless, a limitation of this study is a small data set, which can lower the accuracy of the LDA model, so it should be trained on larger corpora in the future. The lack of similarity between the topics in the corpora needs further qualitative investigation of the speeches to understand how well the discussed topics represent the data.

CONCLUSION

- There is a high Jaccard Distance between the compared topic models from corpus 1 and 2. Therefore, a significant dissimilarity between the topics extracted by the models is indicated. This suggests that the models are decorrelated and may capture distinct aspects or themes within the dataset.
- The Intertopic Distance Map shows that topics in the first speeches are more related than topics in recent speeches of President Zelenskyy.
- The data indicates a shift in topics, reflecting the dynamic nature of war and the ever-changing circumstances that occur on a daily basis.