Group Representations (and other themes) in Soviet Satire from *Krokodil*

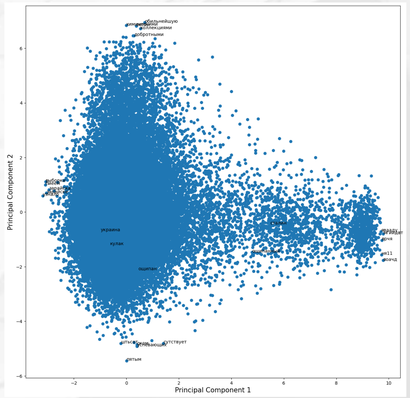
David Axelrod
Department of Informatics, Indiana University

Overview

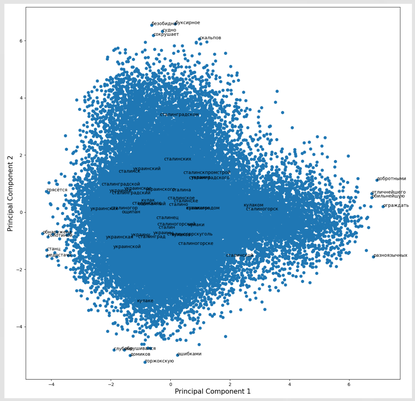
This project explores recurring themes in satirical texts and images from the magazine *Krokodil*, the main satirical publication in the Soviet Union for the period I examine. The corpus was selected in part because satire can provide a revealing medium for analyses of social constructs and biases, not least with respect to group identity. In addition, the *Krokodil* articles of this period reflect perceptions of economic exploitation and progressive economic or political change, which were frequently interpreted along ethnic and urban-rural lines.

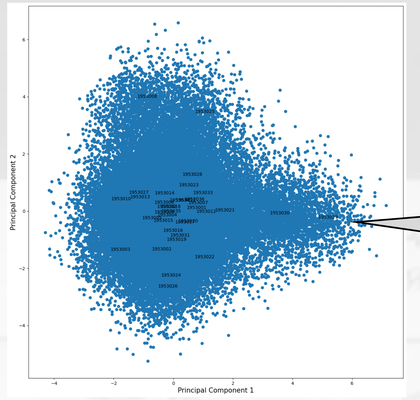
Methods

The first task was to convert PDFs of *Krokodil* issues into machine-readable text and process the text to remove non-relevant symbols by filtering for Cyrillic characters. I then applied several methods for identifying themes or topics in articles from 1953. These include topic modeling with Latent Dirichlet Allocation and clustering texts at various aggregate levels using HDBSCAN. Finally, I created word embeddings using the Word2Vec algorithm and then reduced the embedding space with Principal Component Analysis.



PC Biplot of Word Embeddings. X-Axis Corresponds to Typos Introduced by OCR.





PC Biplot of Word (Top) and Issue (Bottom) Embeddings with Typos Removed.

Findings

Topic modeling at the issue level and clustering texts at issue and sub-issue levels achieved mixed results for finding substantive patterns. Much of the structure picked up in these approaches relates to the format of the publication rather than commonalities in the narratives. This is likely a product of the highly referential nature of the satire found in these articles. One of the more positive findings was the ability to use word embeddings to remove typos from the OCRed text in a scalable manner.

Next Steps

To date, I have limited this proof-of-concept effort to issues from a single year. However, increasing the size of the corpus to cover a greater timespan may help by increasing the likelihood of repeat cultural references. Similarly, word embeddings generally require a larger training corpus and so it would be reasonable to expect increased performance even if the texts were not particularly challenging, as they are. Lastly, it may also prove more meaningful to compare issues across multiple years since one would expect greater variety in the topics being discussed.


ДЕМОНСТРАЦИОННЫЙ ЗАЛ
Крокодила
ПУТЕВОДИТЕЛЬ ПО ЗАЛУ
Детская коляска «ТЯЖЕЛОАТЛЕТ»
Замок «РАДОСТЬ ПРАВОНАРУШИТЕЛЯ»