LINGUISTIC FEATURES OF GERMAN POLITICAL PARTIES

AARON MCKEEVER, COLUMBIA COLLEGE

ADVISING PROFESSOR: DR. ANNIE PFEIFER, COLUMBIA UNIVERSITY

BACKGROUND

- *The rise and fall of the Nazi Germany and the subsequent new republic that was borne from the ashes of WWII has created large bodies of scholarship in field spanning political science, linguistics,*
- *Political science scholars have written about the unorthodox tactics utilized by far-right political parties and the effects of such actions on public trust in the democratic process.*
- *A noticeable change in the political climate was already occurring as a result of economic issues in the EU and the "Refugee Crisis" which peaked in 2015.*
- *These issues have lead to a massive shift in voting, and it resulted in major losses for the previous grand coalition parties (CDU and SDP).*
- *With the rise of social media, politicians have changed the way that they interface with their constituents and the population at large.*
- *Major criticism have been raised by a various of different social research institutions, as well as by other German politicians, about the messages that the AfD uses. Many have been wary about the ties to neo-Nazi ideology.*
- *A key part to my research is from a question laid out by 20th Century scholar Victor Klemperer in his book Language of the Third Reich, which asks whether or not it is possible to get rid of Nazi thoughts if there is not a "denazification" of the language.*
- *Rises in hate speech are followed by rises in hate crimes, and with the concurrent rise of the AfD, their are lot of concerns about the ethicality of stoking fear and inciting violence in politics.*

RESEARCH QUESTIONS

- *Do far-right political parties have statstically unique linguistic features?*
- *Looking at the two different registers, formal being the political documents and informal being social media, are there unique linguistic phenomena that can be understood through the use of natural language processing and computational text analysis?*

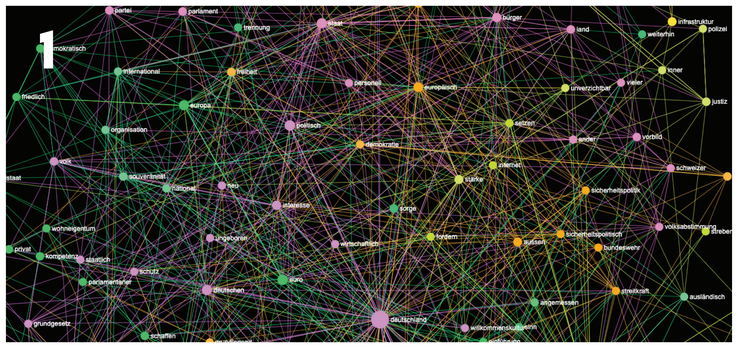
METHODS

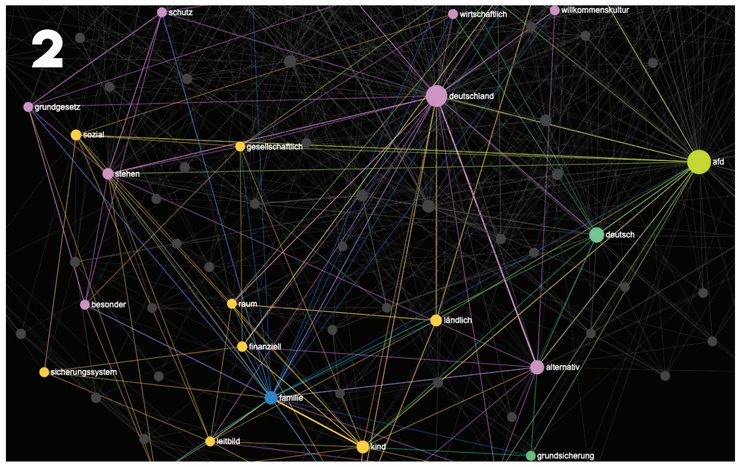
Party Documents (Grundsatzprogramm)

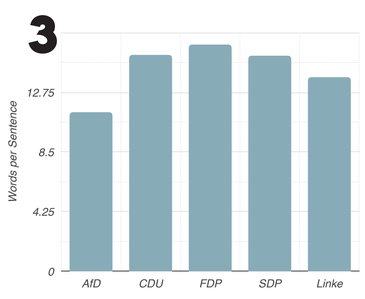
- *The text for the visualizations was pulled from the party documents, so all the text had to be transferred into plain text files and cleaned.*
- *Used Python and the NLP libaries NLTK and Stanfords' CoreNLP to run some initial computations on the corpus.*
- *Used Infranodus, a visualization software that uses graph theory in combination with NLP techniques to create WordNet-like visuals, to glean a lot of initial analyses from these bodies of text. The open source software was able to modified to weed out an extended list of German stop words which made it easier to further isolate the key words.*
- *Tracing specific "lineages" of graphs allowed for comparisons between the different party's most commonly used words in association with important concept words like "justice" and "freedom" (translated).*

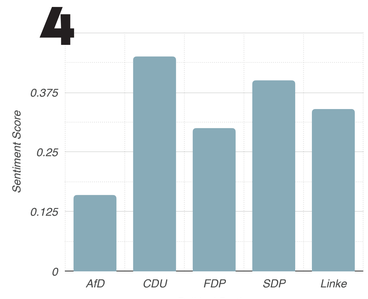
Twitter Data:

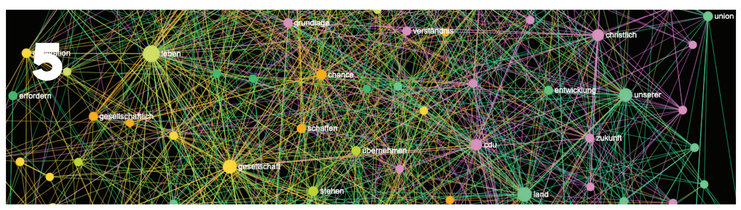
- *The Twitter data used a different approach which utilized more direct implementations of textual analysis using bag of words model.*
- *Using the Twitter developer module, the last 500 tweets of 25 different Bundestag representatives were pulled, 5 from each of the respective parties in the graphs.*
- *This was done using a pseudo-random selection model – using a random number generator over a list of Bundestag representatives. This was done to try and alleviate some sampling biases.*
- *The pulled tweets were run through an algorithm which computed the average word, sentiment score, and most common bi-grams using Python libraries, NLTK, SpaCy and CoreNLP.*











1. *A visualization of the AfD's Grundsatzprogramm using the Infranodus word-network visualization library. This shows the frequency of the the words by the size of the node and the connectivity between words by the edges on the graph. Color reflects a preliminary topical grouping, but refining is needed to correct the model.*

2. *A lineage or selection of the tree for AfD's use of the word "family" (trans.). This shows the connections to other words. In this example, there is a strong appeal to "traditional" (trans.) families and the concept of the AfD "protecting" (trans.) the family unit from cultural shifts. The word "welcome culture" (trans.) is used to describe the way that the AfD will protect the "unborn" (trans.), who are seen as discriminated against.*

3. *A graph describing the average words per sentence in the tweets of each of the five parties. The AfD scores lowest by signficant amount. Interestingly, the opposite extreme, Die Linke, also scored slightly lower. Implementation of an average fluctuation for the error bars in forthcoming in my algortihm.*

4. *Average sentiment graph across the body of tweets. The sentiment was calculated for each individual tweet and then average across and indivudal and then party. The AfD, also scores significantly lower higher. Sentiment scores range from +1 to -1. The closer to +1, the more positive the tweet is, and the closer to -1, the more negative it is. Surpisingly, all parties netted in the positive range.*

5. *A subsection of the CDU shows a completely different vocabulary reflecting "development", "understanding", and "future" (trans.).*

PRELIMINARY RESULTS

Prominent Linguistic Features

Descriptive:

- *Alienation of foreign concepts via "Anglicization", concepts that the party finds to be unacceptable or contrary to their beliefs, the AfD puts them in English and in quotes to set apart as alien or foreign concepts, not indigenous to German culture and thought. (Ex. "Gender-Maintstreamings" (AfD, 109), "Gender-Studies" (AfD, 109), "politischen Korrektheit" (AfD, 114), etc.)*
- *Significant amounts of anti-feminist language is used by the AfD. As listed in the examples above, the AfD spends a large part of their "education" (translated) section rebuking gender, gender studies, and feminist ideas. This section had a disproportionately low sentiment score.*
- *This adherence to "traditional" gender roles is opposite to almost all other parties, especially Die Grüne, who devotes entire subsection to "women politics" (translated).*
- *With the AfD's platform documents being the most recent, it utilizes more heavily currently existing social turmoil, thus seems more immediately relevant. (More information in Further Research.)*

Statistcal:

- *As seen in the graphs, initial research shows the AfD has a few unique characterstics that can be shown through more direct statistical calculation.*
- *The semantic analysis was considerably lower than the other parties, but is net postive which surprised me. However, this is not taking into account the emotional values in attatched images nor is the model using a semantic calculation which takes into account the words place in the order of the sentence.*
- *The initial graphs suggest that more underlying trends exist.*

Interpretation

- *In the AfD there is significant word choices which reflect early Nazi doctrine which calls a prioritzation and unflinching pride in the German identity. Concepts surrounding a return to the traditional culture and appeal to the "traditional family" (trans.) works to delegitimize other ways of life, which is reinforced in the AfD's outright condemnation for women's equality movements.*
- *Twitter usage by parties was more difficult to derive qualitiative data from because of the constant intermixings of retweets or replies which were not neccessarily part of their platform. There were more general differences that could be noticed between the language between the inner parties and outer parties (charged vs. neutral language, "dogwhistling", etc.).*

FURTHER RESEARCH

- *In relation to the Twitter Data, I would like to implement a wider variety of calculations looking at frequency of tweets, audience overlaps, and commonality of language using Zipf's Law calculations. I would also like to visualize the interactions and references of Twitter party members to see if certain parties reference within their own party or reply to others more or less often.*
- *Existing calculations will need to be refined, due to a minor issue with German ordinal system which uses a period to denote place or date (ex. 15. Mai = May 15th, jede 4. Frau = every 4th woman).*
- *With more mathematical knowledge, I hope to be able to implement more robust and insightful statistical models.*
- *With more finetuning of my algorithm, my goal is to run larger sets of data with access to campus high speed computational power. This will improve my statistical signficance and result validity.*
- *Implementation of average variation tracker, to use for standard deviation and standard error calculations.*
- *New software and libraries are always forthcoming. Infranodus is quite new, so I hope to work with it more to create a method for checking overlap between models.*
- *Currently, all parties are finalizing the new versions of their party platforms. Some of the currently used documents are extremely outdated (ex. CDU 2007 vs AfD 2016). These updates will likely cause some significant changes to the result found here, but this new data would be more comparable.*

ACKNOWLEDGEMENTS

- *I want to thank the Laidlaw Foundation and Columbia URF for its support in funding and aiding the research of my project.*
- *I want to thank my advising professor Dr. Annie Pfeifer for helping me to gather my foundational texts, decipher the vocabulary, and refine my corpora.*
- *I also want to thank my graduate student mentor Silvio Ravaioli who helped orient me in the field of social sciences and advised me through the initial stages of my research.*

REFERENCES

- *Klemperer, Victor, 1881-1960. (2002). The language of the Third Reich : LTI, lingua tertii imperii : a philologist's notebook. London ; New York :Continuum,*
- *AfD Grundsatzprogramm. (2020, February 19). Retrieved August 4, 2020, from https://www.afd.de/grundsatzprogramm/*
- *CDU Grundsatzprogramm. Retrieved August 12, 2020, from https://www.cdu.de/partei/dokumente*
- *Twitter Coropra (forthcoming to GitHub)*