

Comparative Analysis of Large Language Models' Stylistic Responses

Maximilian Omlor

March 14, 2024

Contents

1	Introduction	1
2	Research Agenda	2
3	Data overview	4
4	Method overview	6
4.1	Stylometry	6
4.2	Topic Modelling	6
5	Related work	7
6	Experiment Design	8
6.1	Investigate methods for comparing topics	8
6.2	Investigate methods for comparing style	8
6.3	Compare prompt formulations	8
6.4	Compare augmented prompts	9
6.5	Classify ChatGPT speeches between climate change and political figure speeches	9
6.6	Potential difficulties	10
7	Results And Discussion	11
7.1	Investigate methods for comparing topics	11
7.2	Investigate methods for comparing style	11
7.3	Compare prompt formulations	17
7.4	Compare augmented prompts	22
7.5	Classify ChatGPT speeches between climate change and political figure speeches	22
8	Conclusion	23
9	AI Tool Usage	25
9.1	Code generation for a problem description	25
9.2	Error search by specifying code and error.	27
9.3	Feature expansion by providing the previous code passage and a verbal description of a desired feature.	27
9.4	Translating text passages.	27
9.5	Generating a paragraph	28
9.6	Summaries of texts or web pages	29
9.7	Structural recommendations, such as the organization of a chapter	30

List of Tables

1	Overview of all granular research questions	2
2	Overview of the various data sources with number of samples	4
3	Overview of content generated with ChatGPT with number of samples	5
4	Overview of the various Prompts used to generate Speeches with ChatGPT 3.5 and 4 . . .	9
5	Overview of the various augmented Prompts used to generate Speeches with ChatGPT 3.5 and 4. used speeches by Date ID are 1904-02-19, 1920-11-04 and 1945-05-08	9
6	Prompt Comparison Evaluation on Topic by eye. 1st to 4th describes which sample is closest to the topic cluster. (in) indicates that the sample has landed in the topic cluster . .	19
7	Prompt Comparison Evaluation on Style by eye. 1st to 4th describes which sample is closest to the style cluster. (in) indicates that the sample has landed in the style cluster . .	21

List of Figures

1	Descriptive statistics over the used data sets	4
2	Descriptive statistics over the used GPT datasets	5
3	PCA (corr.) on Bag of Words feature with varying MFW Min and Max.	12
4	PCA (corr.) on Bag of Words feature with varying MFW Min and Max and varying Start at Freq. rank (SAFR).	13
5	PCA (corr.) on Topic Words feature with varying MFW Min and Max.	14
6	PCA (corr.) on Topic Modeling feature with varying MFW Min and Max.	15
7	PCA (corr.) on Character 3-Gram feature with varying MFW Min and Max.	16
8	PCA (corr.) on Sentence and Punctuation Features	17
9	Topic Modelling on cc, jfk and all gpt samples related to jfk	18
10	Topic Modelling on cc, mt and all gpt samples related to mt	18
11	Topic Modelling on cc, wc and all gpt samples related to wc. gpt35ragwc (black) and gpt4ragwc (brown) are prompts augmented with original Speeches from Winston Churchill	19
12	PCA (corr.) on Feature Char 3-Gram on cc, jfk and all gpt samples related to jfk	20
13	PCA (corr.) on Feature Char 3-Gram on cc, mt and all gpt samples related to mt	20
14	PCA (corr.) on Feature Char 3-Gram on cc, wc and all gpt samples related to wc. gpt35ragwc (black) and gpt4ragwc (brown) are prompts augmented with original Speeches from Winston Churchill	21

1 Introduction

The progressive development of language models based on artificial intelligence has led to remarkable advancements in recent years, opening up new possibilities for text generation. Particularly, advanced language models like ChatGPT have the potential to respond to complex queries and produce high-quality texts. In various contexts, the question arises as to whether these language models are capable of imitating existing texts in their topics and styles, and whether these texts can be distinguished from their originals.

This project originated within the scope of the "Introduction to Digital Humanities" module. But what does Digital Humanities stand for, and what is the significance of Digital Humanities?

According to [1], Digital Humanities (DH) combines the humanities with computer science, focusing not only on language and literary studies but on all humanities and cultural studies. DH utilize digital technologies for digitization, archiving, representation, visualization, and analysis of humanities data. While traditional humanities often conduct manual analyses, DH enable automated procedures and deep insights through the processing of large datasets.

This project thus aligns perfectly with the field of Digital Humanities through the analysis of political speeches, texts generated by ChatGPT, and methodologies such as stylometry. It falls predominantly within the realm of Computational Linguistics, as it focuses less on the content of the texts and more on the structure and style of the texts being analyzed and compared.

Within this project, we investigate how an advanced language model like ChatGPT behaves when tasked with generating speeches on contemporary topics such as climate change. The focus lies on whether ChatGPT is capable of producing speeches in the style of well-known historical figures like Winston Churchill, Margaret Thatcher, and John F. Kennedy. Additionally, we analyze whether stylometric methods can identify the differences between the texts generated by ChatGPT and authentic speeches by these personalities, as well as real speeches on the topic of climate change.

The findings from this investigation are highly relevant as they not only provide insights into the performance of language models like ChatGPT but also shed light on the application of stylometric analyses for evaluating AI-generated texts.

This project report is divided into several sections. First, the research agenda is presented. Subsequently, the utilized data sources and methodologies are described. Following this, an overview of the current state of research in this field is provided. The experiment design is then introduced. Afterward, the results and analysis are presented, and their implications are discussed. Finally, conclusions are drawn, and a perspective on potential future research directions is provided.

2 Research Agenda

The research question investigated in this project is as follows:

How does the advanced language model ChatGPT respond to requests for generating speeches on a modern topic like climate change, in the style of historical figures like Winston Churchill, Margaret Thatcher, and John F. Kennedy, and can stylometric analyses distinguish the differences between these AI-generated texts and authentic speeches of these figures, as well as real speeches on the topic of climate change?

The idea behind this question is to task the language model ChatGPT with generating a text based on a combination of a topic and a person that does not exist in reality, such as Winston Churchill and climate change. The assumption is that this approach can rule out the possibility that ChatGPT was trained on a corresponding text, allowing the generated text to make a statement about the style and ability of ChatGPT to imitate texts and personalities. Original speeches by Winston Churchill and authentic speeches on the topic of climate change will be used as reference points for the analysis of the generated texts.

From the research question, several more granular questions arise, which are listed again for better clarity in Table 1:

	Question
1	Investigate methods for comparing topics
2	Investigate methods for comparing style
3	Compare prompt formulations
4	Compare augmented prompts
5	Classify ChatGPT speeches between climate change and political figure speeches

Table 1: Overview of all granular research questions

Two different comparison approaches are needed. On one hand, an approach is required to compare the themes of the speeches. This is necessary to determine how similar the speeches generated by ChatGPT are to those about climate change. On the other hand, another approach is needed to examine the style of the speeches independent of the topic. This is required to assess the similarity between the ChatGPT speeches and those of historical figures. Therefore, a few possible approaches should be investigated and compared.

For targeted text generation with ChatGPT, the formulation of the prompt is essential. Therefore, several different prompt formulations should be tested and compared. Additionally, it should be investigated whether text generation can be improved if ChatGPT is provided with additional information in the form of an original speech by a historical figure.

Finally, it should be investigated how the speeches generated by ChatGPT compare to both the speeches of historical figures and those about climate change. Can the ChatGPT speeches be clearly distinguished from the others, or can ChatGPT so accurately imitate the themes and styles that they are indistinguishable using stylometric analysis?

The data used to address these questions are derived from speeches by historical politicians Winston Churchill, Margaret Thatcher, and John F. Kennedy, speeches about climate change by Antonio Guterres,

Greta Thunberg, and various contributions from leading politicians at the United Nations Framework Convention on Climate Change (UNFCCC), as well as texts generated with ChatGPT 3.5 and ChatGPT 4.

For analysis, stylometric methods such as Principal Component Analysis (PCA) on Bag of Words, character 3-grams, sentence length, punctuation, and Bag of Words on a previously conducted topic modeling will be used.

3 Data overview

Due to the lack of a suitable dataset that connects speeches from different politicians with speeches from various individuals on the topic of climate change, an additional master's project effort was dedicated to compiling a dataset. This dataset comprises speeches from different sources, categorized into speeches by historical political figures and speeches about climate change. In cases where interview situations occurred in the speeches, these interview segments or irrelevant individuals for this investigation were removed, as only the text of the relevant individuals is pertinent to the research question.

Acronym	Cathegory	Description	Scource	Samples
jfk	historical political figure	speeches of John F. Kennedy	[2]	100
mt	historical political figure	speeches of Margaret Thatcher	[3]	10
wc	historical political figure	speeches of Winston Churchill	[4]	50
un	climate change	speeches of leaders around the United Nations	[5]	28
gt	climate change	speeches of Greta Thunberg	[6]	12
unfccc	climate change	speeches of COP 26 UNFCCC	[7]	35

Table 2: Overview of the various data sources with number of samples

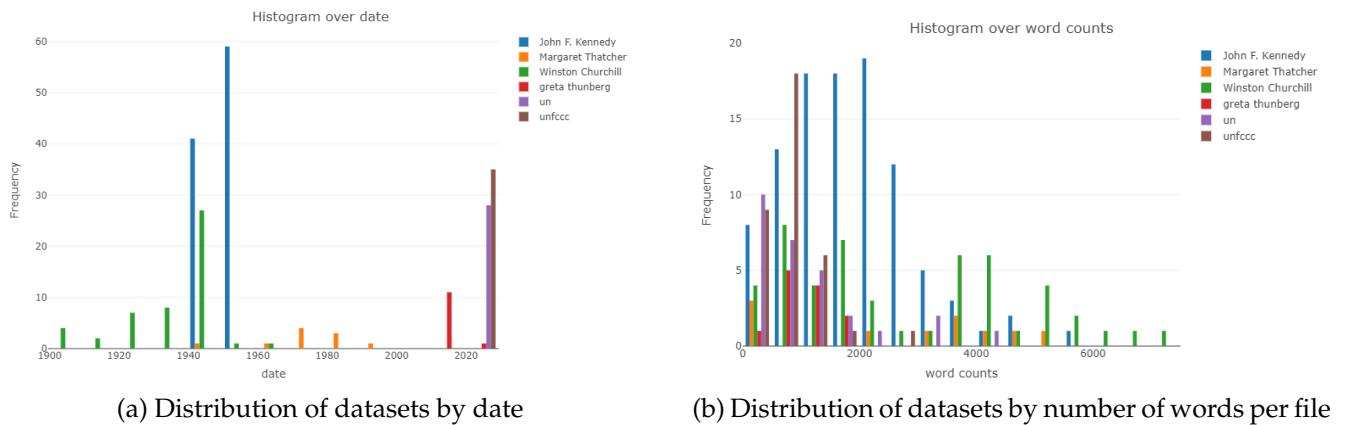


Figure 1: Descriptive statistics over the used data sets

The various datasets vary in size, as evident in the Sample column of Table 3. The assumption is that this is not overly problematic, as the method used can handle datasets of different sizes. A significant temporal discrepancy is observed in the data (see Figure 1a). The speeches by historical politicians are clearly positioned earlier in time than those about climate change. This is expected due to the experiment design but could potentially lead to challenges in comparability of results due to linguistic changes over time. The length of the speeches, indicated by the number of words (see Figure 1b), also varies greatly. Depending on the speakers and the contexts in which the speeches were delivered, they can be relatively long or very short. Generally, the statistical power of a stylometric analysis, for example using the Bag-of-Words method, increases with longer text samples. Consequently, the UNFCCC samples should be approached with more caution compared to, for example, the Winston Churchill samples. It is also worth mentioning that all UN samples consist of speeches by one person, namely Antonio Guterres, while the UNFCCC samples are all attributed to different individuals. Therefore, greater similarity is expected among the UN samples compared to the UNFCCC samples.

Additionally, a dataset of generated speeches using ChatGPT was created, utilizing both versions ChatGPT 3.5 and ChatGPT 4. In cases where ChatGPT generated introductory and/or concluding descriptive text, only the speech part was used for analysis.

Acronym	Cathegory	Description	Samples
gpt35_cc	ChatGPT 3.5	speeches on climate change	3
gpt35_ccjfk	ChatGPT 3.5	speeches on climate change in style of John F. Kennedy	4
gpt35_ccmt	ChatGPT 3.5	speeches on climate change in style of Margaret Thatcher	4
gpt35_ccwc	ChatGPT 3.5	speeches on climate change in style of Winston Churchill	4
gpt35rag_ccwc	ChatGPT 3.5	speeches on climate change in style of Winston Churchill augmented by original speech	9
gpt35_jfk	ChatGPT 3.5	speeches in style of John F. Kennedy	4
gpt35_mt	ChatGPT 3.5	speeches in style of Margaret Thatcher	4
gpt35_wc	ChatGPT 3.5	in style of Winston Churchill	4
gpt4_cc	ChatGPT 4	speeches on climate change	3
gpt4_ccjfk	ChatGPT 4	speeches on climate change in style of John F. Kennedy	4
gpt4_ccmt	ChatGPT 4	speeches on climate change in style of Margaret Thatcher	4
gpt4_ccwc	ChatGPT 4	speeches on climate change in style of Winston Churchill	4
gpt4rag_ccwc	ChatGPT 4	speeches on climate change in style of Winston Churchill augmented by original speech	9
gpt4_jfk	ChatGPT 4	speeches in style of John F. Kennedy	4
gpt4_mt	ChatGPT 4	speeches in style of Margaret Thatcher	4
gpt4_wc	ChatGPT 4	in style of Winston Churchill	4

Table 3: Overview of content generated with ChatGPT with number of samples

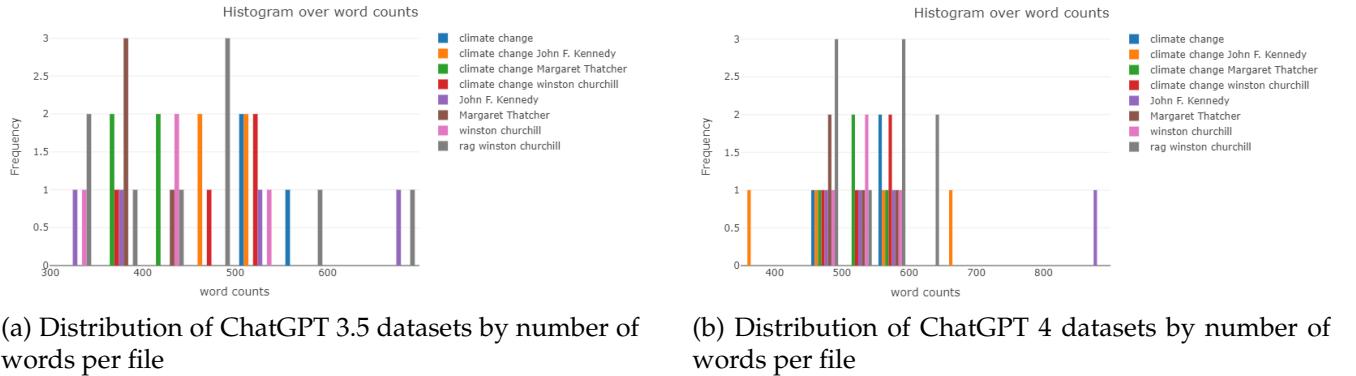


Figure 2: Descriptive statistics over the used GPT datasets

The samples generated by ChatGPT are often considerably shorter than the real-world samples (see Figure 2).

For specific analyses, subsamples of the entire dataset were created to form smaller corpora. For example, this was done for analyses focused on individual historical figures. Further details on the corpora will be provided in sections 6 and 7.

4 Method overview

In this project, several methods of stylometry and topic modeling were used and combined. The following sections will elucidate the methods employed.

4.1 Stylometry

Stylometry is generally used to identify similarities in texts [8]. To achieve this, various quantifiable textual features can be analyzed, such as the words contained in the text, sentence length, frequencies of (groups of) letters and (groups of) words, frequencies of grammatical constructions, frequencies of rare words, and many others. From these analyses, feature vectors can be derived to describe a given text based on its features. These vectors can then be compared using various distance measures or through the representation of multiple vectors in a PCA (Principal Component Analysis), for example, to make statements about the stylistic similarity of the given texts.

In this project, various features were used, including Bag of Words, character 3-grams, sentence length, punctuation, and some features associated with topic modeling (which will be discussed further in that section).

Bag of Words describes a method that counts the occurrences of each individual word in a given text and summarizes these counts in a vector. Character 3-grams count the occurrences of sequences of 3 consecutive letters instead of whole words. These counts are also summarized in a vector.

Difficulties that may arise in using these methods in relation to the research question include the complexity and subtle differences that can define a style. Representing these through the given methods and interpreting the results accordingly could prove challenging.

4.2 Topic Modelling

Topic modeling is necessary to determine the semantic meaning of words [9]. Since there are words that can have multiple meanings depending on the context, a simple frequency analysis is not sufficient to make a statement about their semantic meaning. For this purpose, topic models are employed. These model topics as lists of words that statistically co-occur frequently in a given text corpus.

In this project, topic models were used to generate features for a frequency analysis.

5 Related work

Stylometry has historically been used in a variety of linguistic application areas. Below is a brief overview of some published papers in this field.

Song Feng et al. explore the application of stylometry in deception detection in their paper titled "Syntactic Stylometry for Deception Detection" [10]. They utilize features derived from trees parsed by a Context-Free Grammar. Through this approach, they were able to significantly improve upon the best-published results in deception detection.

Glory O. Adebayo and Roman V. Yampolskiy reviewed studies on IQ estimation and author profiling with stylometry in their paper "Estimating Intelligence Quotient Using Stylometry and Machine Learning Techniques: A Review" [11]. They conclude that, given a good dataset, the use of stylometry in a study on IQ estimation from written text promises to yield good results.

In their paper "De-anonymizing Programmers via Code Stylometry" [12], Aylin Caliskan-Islam et al. utilized stylometry to determine the authorship of C and C++ code samples. They derived their feature set from abstract syntax trees and employed a random forest approach. Their method was successfully tested on the Google Code Jam dataset.

6 Experiment Design

The following outlines the methods applied to the data to address the various granular research questions.

6.1 Investigate methods for comparing topics

PCA (corr.) on Bag of Words features with varying Minimum and Maximum Document Frequency (MFW) as well as varying Start at Frequency Rank (SAFR), assuming that topic-defining words occur less frequently than stopwords.

PCA (corr.) on Bag of Words features on text files from which all stop words and punctuation have been removed, assuming that primarily the topic-defining words remain. Hereinafter abbreviated as Topic Words.

PCA (corr.) on Bag of Words features on transformed text files. In this case, topic modeling was applied to each individual text with k equal to 5, and subsequently, the top 20 words of each topic were compiled as text files into the used corpus. Hereinafter abbreviated as Topic Modeling.

All methods were applied to a subset of the dataset, which includes all speeches except for those generated by ChatGPT. It is expected that a method primarily comparing topics will exhibit a greater dispersion of speeches from individual persons. In contrast, different individuals may overlap more strongly, as they can speak about similar topics. A clear difference should be observed between the speeches about climate change and those of the historical politicians, as the assumption is that these are thematically distinct from each other.

6.2 Investigate methods for comparing style

PCA (corr.) on 3-Gram character features, assuming that 3-Gram characters convey more about the style and structure of the text than Bag of Words.

PCA (corr.) on Bag of Words features on transformed text files. Here, the length of the sentences, measured by the number of words in a sentence, is encoded in text files. For this purpose, each sentence is replaced by an artificial word consisting of as many "x"s as the sentence contains words. Hereinafter abbreviated as Sentence.

PCA (corr.) on Bag of Words features on transformed text files. In this case, only punctuation marks were retained and encoded as spelled-out words in the text files. Hereinafter abbreviated as Punctuation.

All methods were applied to a subset of the dataset, which includes all speeches except for those generated by ChatGPT. It is expected that a method primarily comparing style can distinguish between all individuals. This should lead to a relatively large dispersion within the UNFCCC subset, as speeches from very different politicians are included here.

6.3 Compare prompt formulations

Here, several different prompt formulations were tested and compared with the climate change speeches using the most successful topic comparison method. They were then stylistically compared with the speeches of political personalities.

Cathegory	ID	Prompt
cc	01	write me a speech about climate change
	02	write me a long speech about climate change
	03	write me a long and emotional speech about climate change
cc_jfk, cc_mt, cc_wc	01	write me a speech about climate change in style of [Person]
	02	write me a long speech about climate change in style of [Person]
	03	write me a long and emotional speech about climate change in style of [Person]
	04	write a speech about climate change. this speech should sound and feel as if [Person] had written it. pay attention to the sentence structure and style of [Person]
jfk, mt, wc	01	write me a speech in style of [Person]
	02	write me a long speech in style of [Person]
	03	write me a long and emotional speech in style of [Person]
	04	write a speech. this speech should sound and feel as if [Person] had written it. pay attention to the sentence structure and style of [Person]

Table 4: Overview of the various Prompts used to generate Speeches with ChatGPT 3.5 and 4

The structure of the prompts corresponds to the specifications outlined in Table 4.

6.4 Compare augmented prompts

Several prompt formulations augmented with additional information in the form of original speeches by Winston Churchill were also tested and compared. Additionally, a comparison was made to the standard prompts. The expectation is that the augmented prompts will perform better than the simple prompts.

IDs	Prompt	inspired by
01, 04, 07	read this speech of winston churchill: [speech] and write a speech about climate change in style of winston churchill	
02, 05, 08	read this speech of winston churchill: [speech] and write a similar speech about climate change in style of winston churchill	
03, 06, 09	write a speech about climate change in style of winston churchill. use the original speech of winston churchill, written below as a reference. original speech: """[speech]"""	[13]

Table 5: Overview of the various augmented Prompts used to generate Speeches with ChatGPT 3.5 and 4. used speeches by Date ID are 1904-02-19, 1920-11-04 and 1945-05-08

The structure of the prompts corresponds to the specifications outlined in Table 5.

6.5 Classify ChatGPT speeches between climate change and political figure speeches

Comparison of the most successful prompt formulations with the climate change speeches and the speeches of political personalities using the most successful comparison method. It is expected that the samples generated by ChatGPT will land in the respective topic or style clusters depending on the

analysis.

6.6 Potential difficulties

Throughout the project, several issues have arisen, which have led to some of the granular research questions and the performance tests of different methods. Below are some of them examined in more detail.

Firstly, it should be noted that both the number of examined individuals in the historical figures and climate change topics, as well as the number of generated ChatGPT texts, were relatively small. This is mainly due to the given time frame and the learning curve associated with working in the R environment. Therefore, the results obtained do not have general validity and rather serve as a small insight into the capabilities or weaknesses of ChatGPT.

Upon reviewing the initial results, it became apparent that the results of PCA (corr.) on the Bag of Words feature were not only influenced by style but also strongly influenced by the topic (see Section 7). Accordingly, different methods for comparing style and topic were implemented and tested.

At the time of this project, it is unclear to what extent speeches on climate change and speeches by the investigated political figures were included in the training data of the various ChatGPT versions. Furthermore, it is unclear to what extent the temporal bias of the different types of speeches studied influences the performance of ChatGPT. For example, one assumption that arose during the project was that the English language has changed stylistically significantly over the last 100 years, and that ChatGPT may have been primarily trained on more recent texts, thus potentially struggling to imitate historical personalities. This assumption could neither be proven nor disproven within the project scope but poses a challenge in interpreting the results.

For the speeches of different individuals, it is not clear whether they wrote their speeches themselves or to what extent they contributed to them, nor is it known whether they had changing ghostwriters throughout their lives. This could significantly blur the style of speeches attributed to a person.

Speeches, generally speaking, are not particularly long compared to other texts. This excludes, for example, PCA (corr.) over n-grams of words and could potentially yield less accurate results compared to, for example, comparing book texts.

7 Results And Discussion

7.1 Investigate methods for comparing topics

All variants of PCA (corr.) on the Bag of Words feature (see Figure 3) can effectively separate the climate change speeches (red) from the rest of the speeches. In 3c and 3d, a clearer distinction is evident. Some dispersion within individuals is observable. For instance, wc, jfk, and mt overlap to some extent. In contrast, cc_gt is clearly separated from cc_un and cc_unfccc. Surprisingly, a subset of mt aligns closely with cc_gt in all variants.

The variants of PCA (corr.) on the Bag of Words feature with SAFR (see Figure 4) also separate the climate change speeches (red) from the rest of the speeches. There is again an overlap between jfk, wc, and mt, and cc_gt is somewhat separated from cc_un and cc_unfccc. In 4a, it is possible to separate the mt samples from the cc_gt samples. In 4c and 4d, the distance between cc_gt and jfk, wc, mt is relatively small.

The variants of PCA (corr.) on the Topic Words feature (see Figure 5) present a comparable picture to Figure 4. 5a also manages to separate mt from cc_gt. The distance between cc and the rest is consistently evident.

The variants of PCA (corr.) on the Topic Modeling feature (see Figure 6) distinctly separate cc from the rest, although 6a and 6b exhibit some overlap. cc_gt, cc_un, and cc_unfccc show significant overlap in 6a and 6b. mt aligns closer to wc and jfk in 6a, 6b, and 6c.

For comparing topics, based on this analysis, PCA (corr.) on the Topic Modeling feature (see Figure 6) is preferred. This is because of the overlap between cc_gt, cc_un, and cc_unfccc, and the relatively good distinctiveness of mt from cc. For MFW Min Max, the value 100 will be used. Although the values 100 and 300 exhibit very similar characteristics, PC1 with 8.7% and PC2 with 4.5% of the value 100 describe significantly more variance than PC1 (4.5%) and PC2 (2.9%) of the value 300.

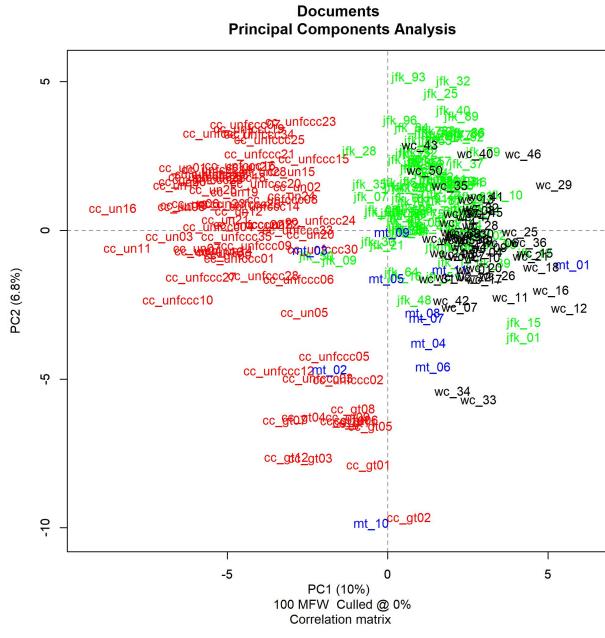
7.2 Investigate methods for comparing style

All variants of PCA (corr.) on the Character 3-Gram feature (see Figure 7) can effectively separate cc_gt from cc_un and cc_unfccc. wc and jfk exhibit significant overlap in all variants. mt shows considerable dispersion and cannot be clearly distinguished from the rest. cc_unfccc shows relatively high dispersion. The individual variants differ little apart from the proportion of variance they describe.

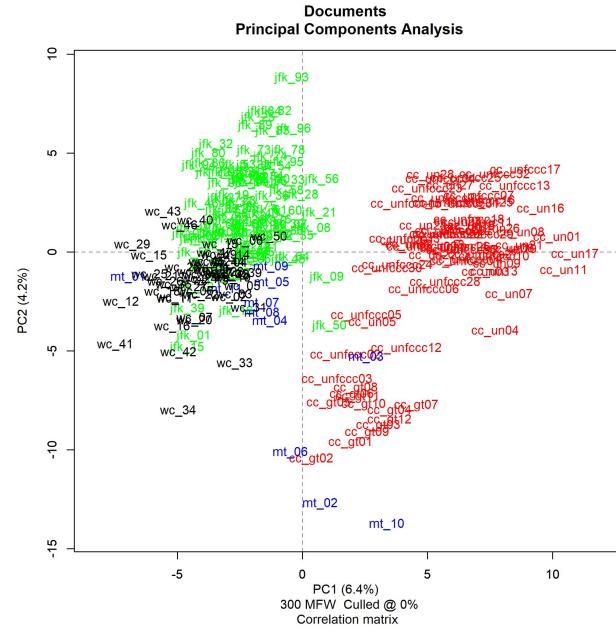
PCA (corr.) on the Sentence feature (see Figure 8a) exhibits significant dispersion across all subsets. jfk, mt, and cc overlap considerably. wc and jfk overlap almost entirely. Individual persons are practically poorly to non-distinguishable. The dispersion of jfk appears significantly larger than that of cc_unfccc.

PCA (corr.) on the Punctuation feature (see Figure 8b) shows a large cluster comprising all subsets, with individual outliers from different subsets. Individual persons cannot be distinguished.

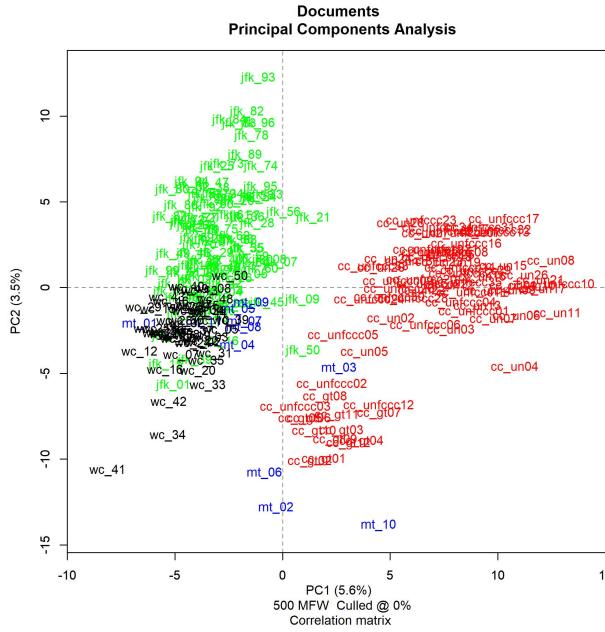
For comparing styles, PCA (corr.) on the Char 3-Gram feature (see Figure 7) is recommended based on this analysis. This is primarily because of the poor to absent differentiation of individual persons in 8a and 8b. For MFW Min Max, the value 100 (7a) is used here, as it describes the largest portion of data variance with PC1 15.4% and PC2 9.6%. The results from 7 deviate significantly from the expected



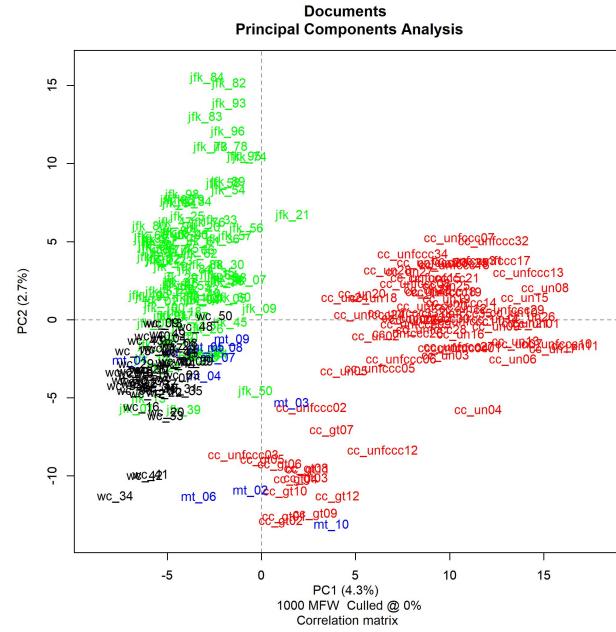
(a) MFW Min and Max equal to 100



(b) MFW Min and Max equal to 300

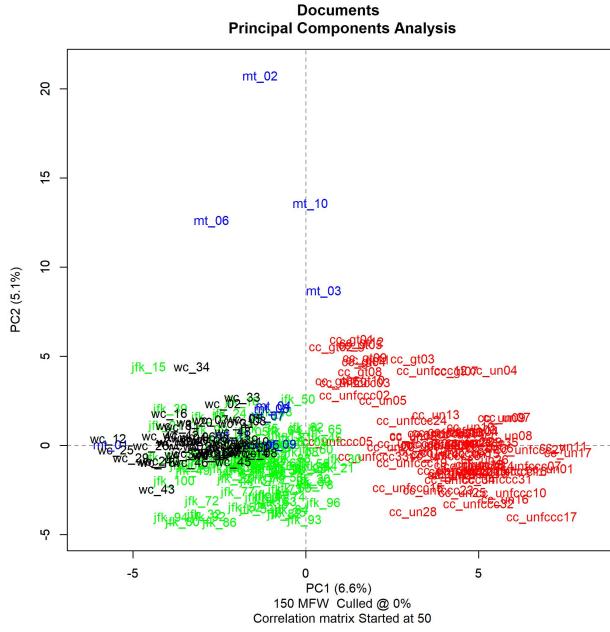


(c) MFW Min and Max equal to 500

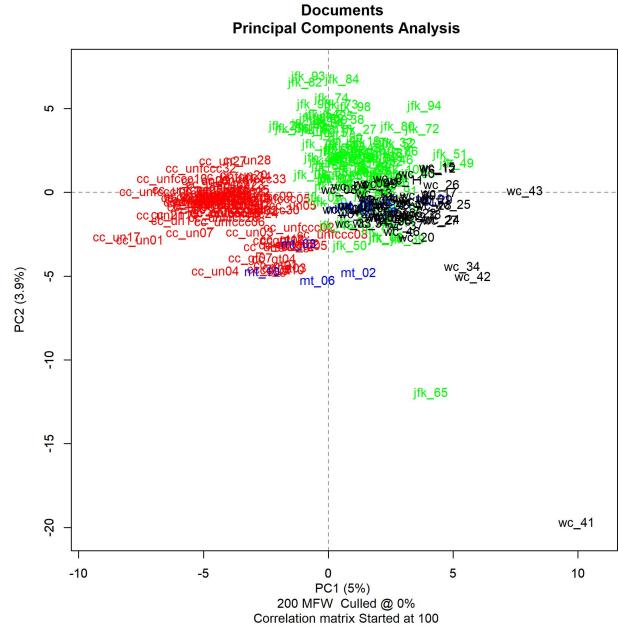


(d) MFW Min and Max equal to 1000

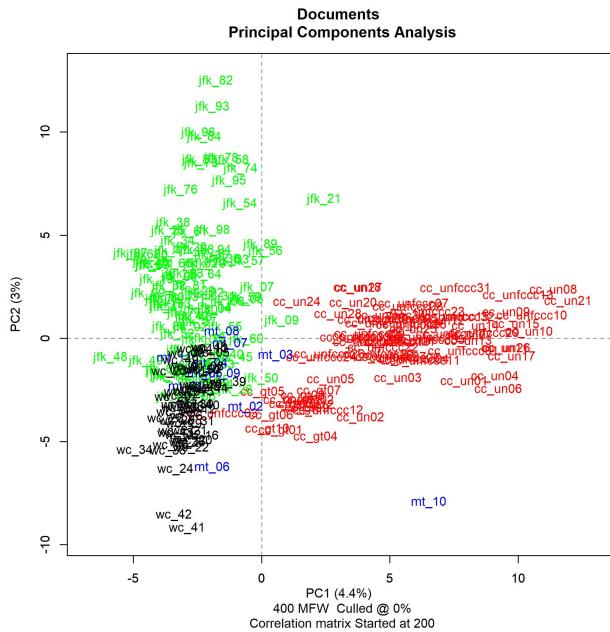
Figure 3: PCA (corr.) on Bag of Words feature with varying MFW Min and Max.



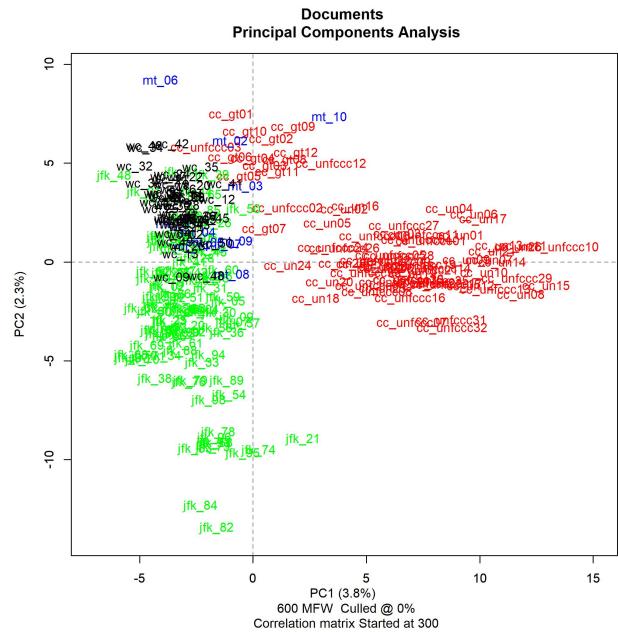
(a) MFW Min and Max equal to 150 and SAFR equal to 50.



(b) MFW Min and Max equal to 200 and SAFR equal to 100.

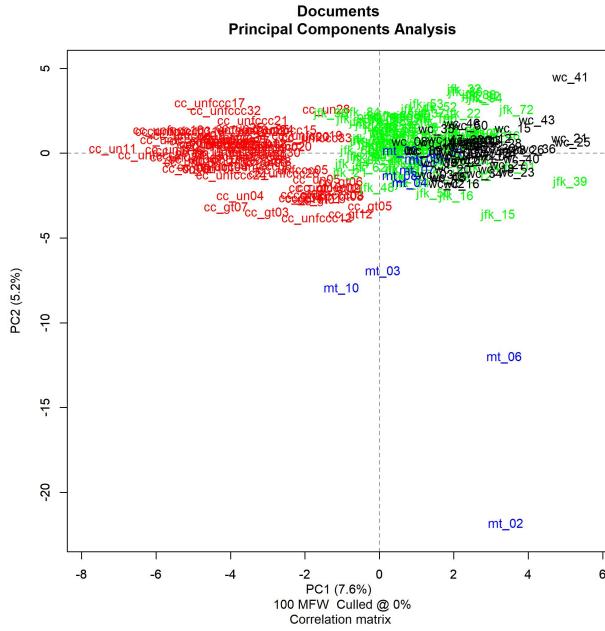


(c) MFW Min and Max equal to 400 and SAFR equal to 200.

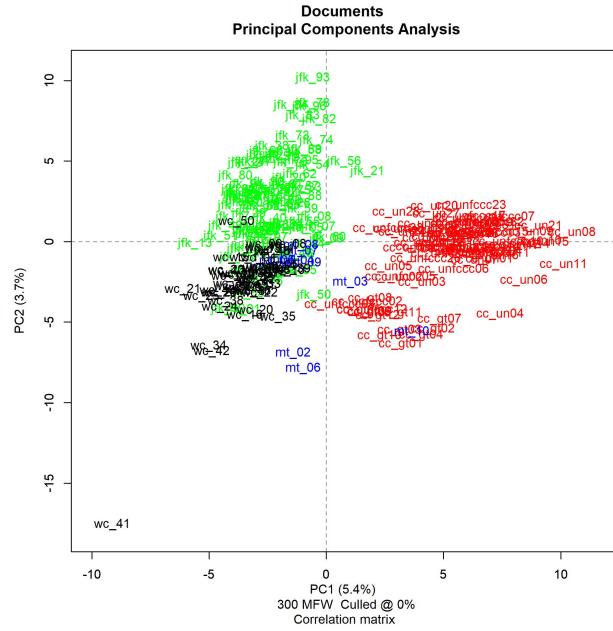


(d) MFW Min and Max equal to 600 and SAFR equal to 300.

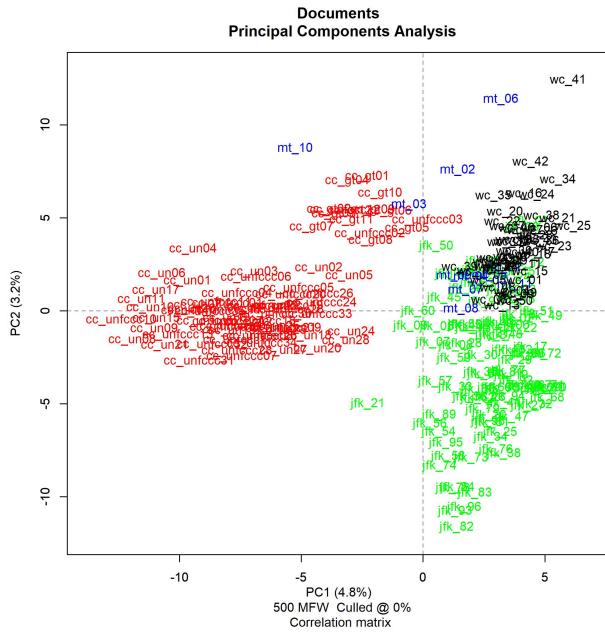
Figure 4: PCA (corr.) on Bag of Words feature with varying MFW Min and Max and varying Start at Freq. rank (SAFR).



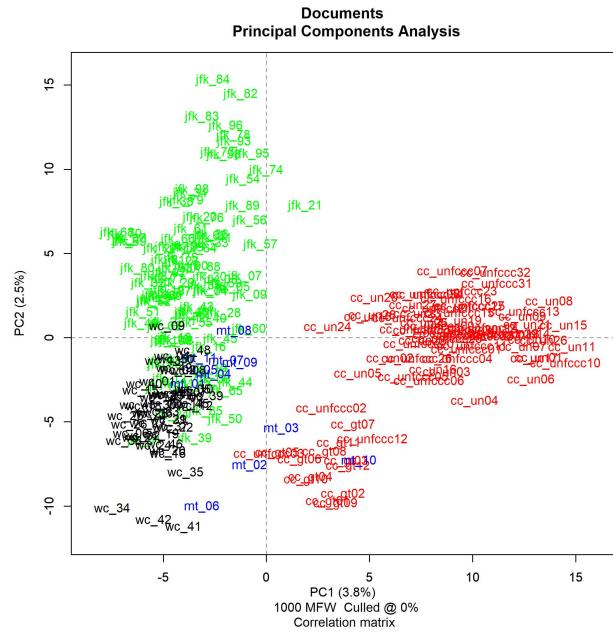
(a) MFW Min and Max equal to 100



(b) MFW Min and Max equal to 300

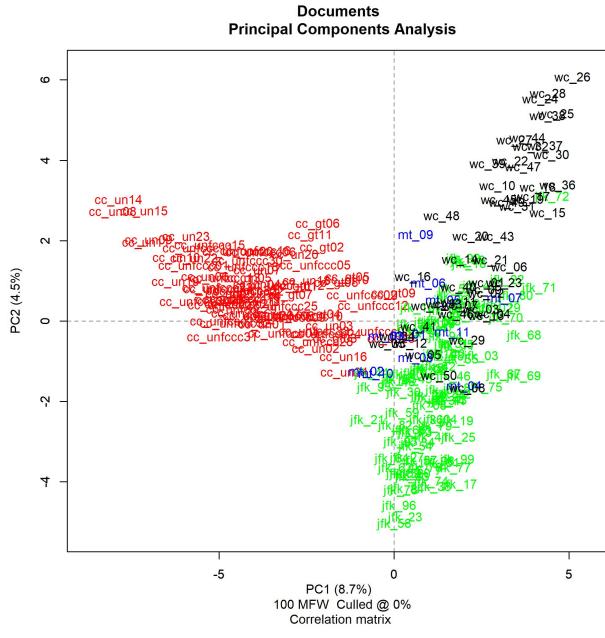


(c) MFW Min and Max equal to 500

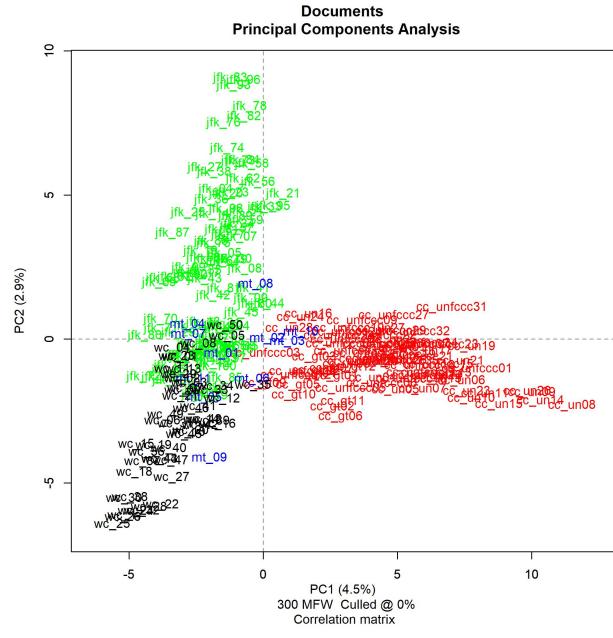


(d) MFW Min and Max equal to 1000

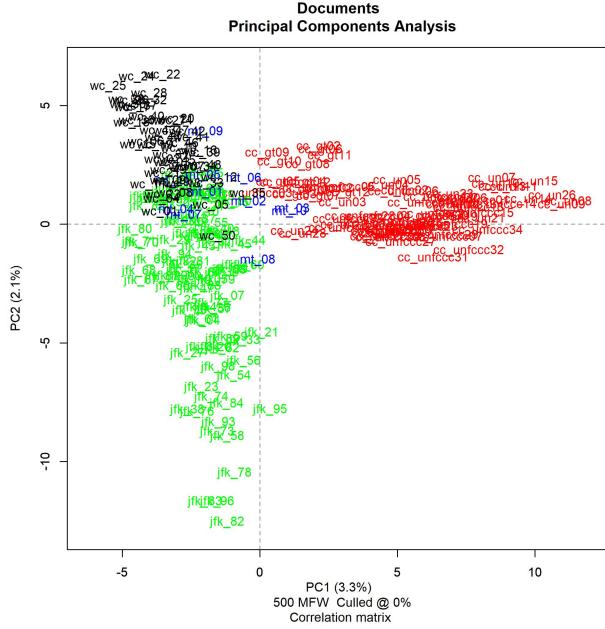
Figure 5: PCA (corr.) on Topic Words feature with varying MFW Min and Max.



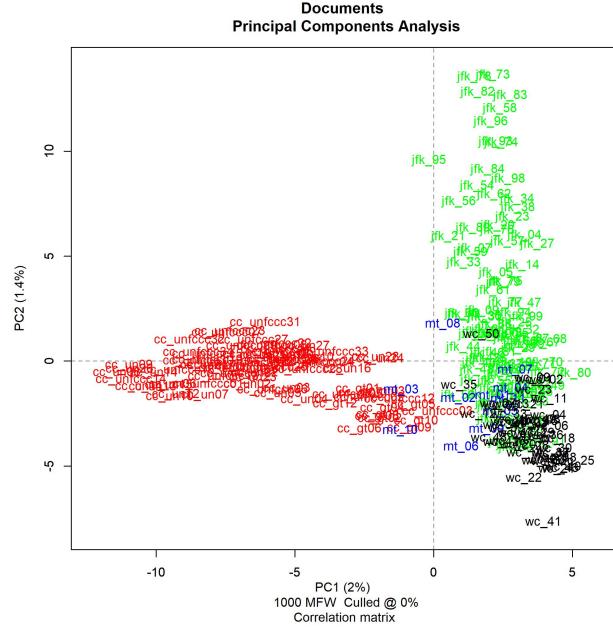
(a) MFW Min and Max equal to 100



(b) MFW Min and Max equal to 300

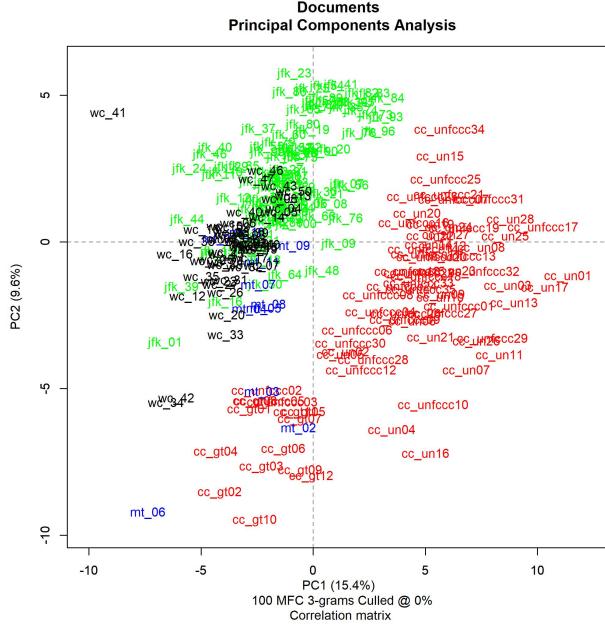


(c) MFW Min and Max equal to 500

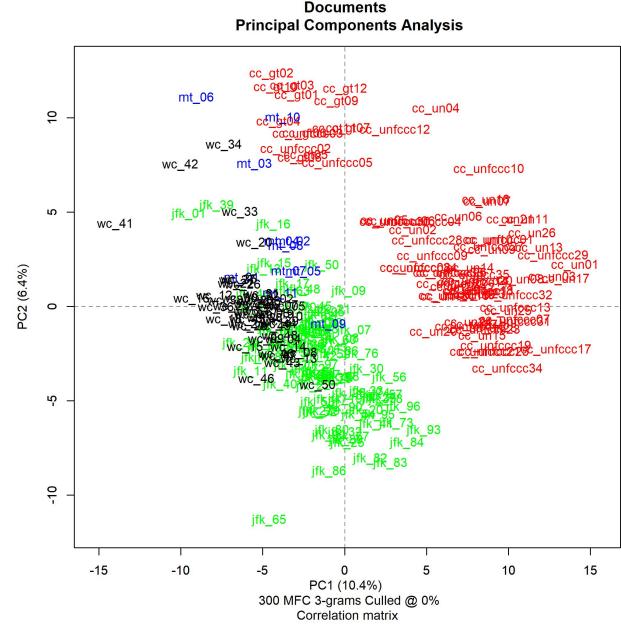


(d) MFW Min and Max equal to 1000

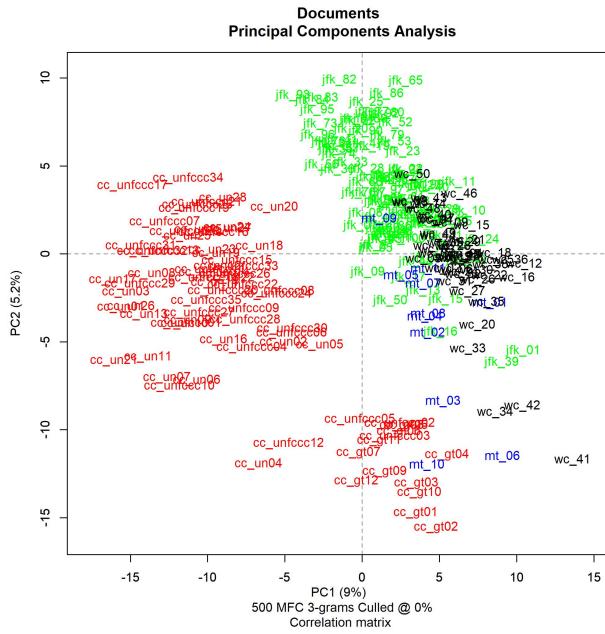
Figure 6: PCA (corr.) on Topic Modeling feature with varying MFW Min and Max.



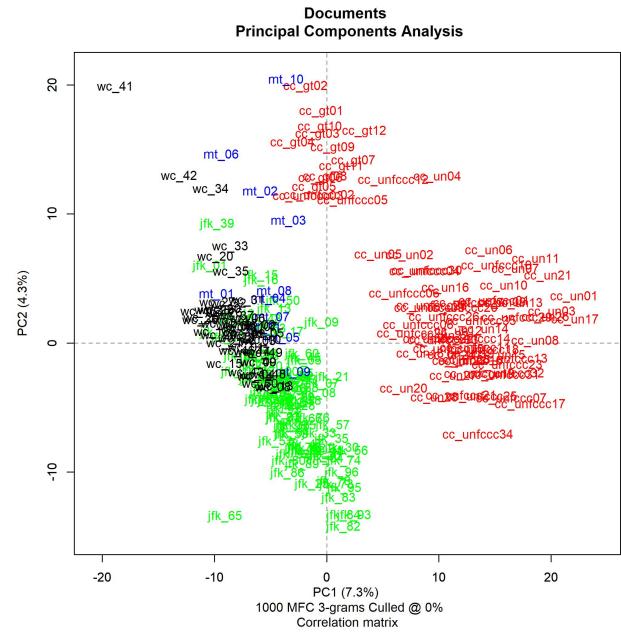
(a) MFW Min and Max equal to 100



(b) MFW Min and Max equal to 300



(c) MFW Min and Max equal to 500



(d) MFW Min and Max equal to 1000

Figure 7: PCA (corr.) on Character 3-Gram feature with varying MFW Min and Max.

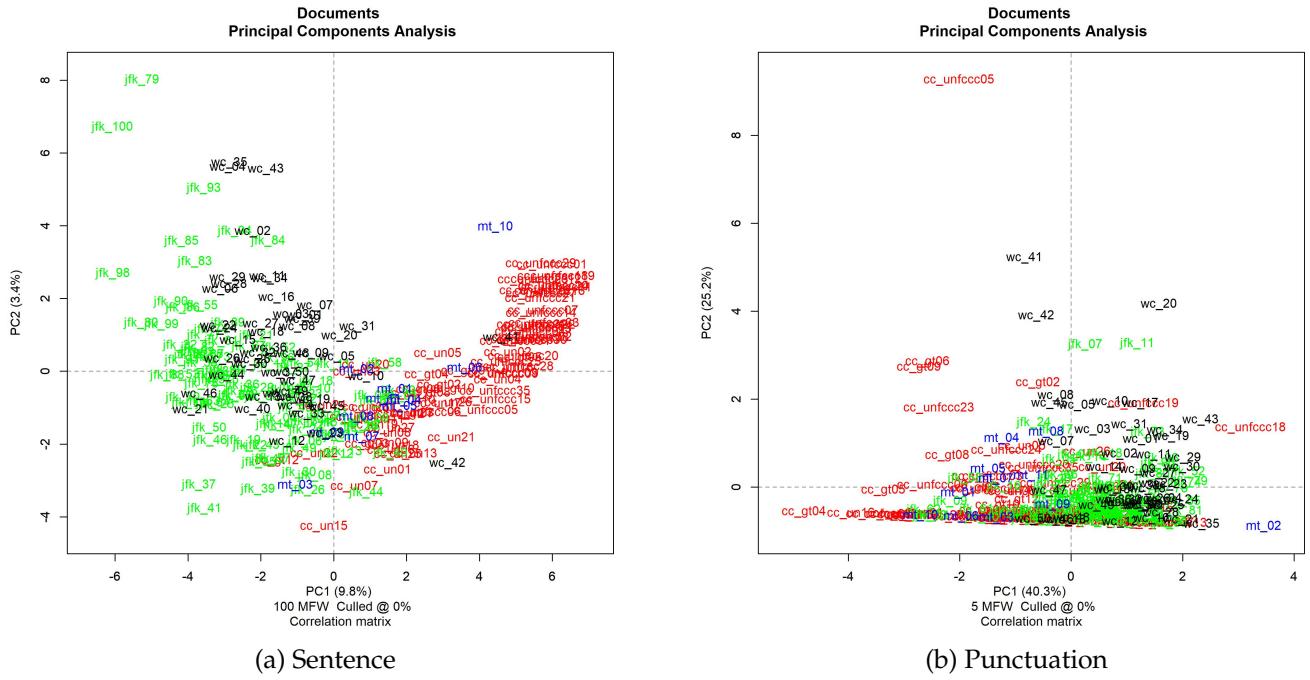


Figure 8: PCA (corr.) on Sentence and Punctuation Features

outcome. Therefore, as a future outlook, it would be advisable to further investigate the data and explore additional style analysis methods.

7.3 Compare prompt formulations

Figures 9, 10, and 11 depict the comparison of the various tested prompts analyzed by topics. Table 6 summarizes the resulting outcomes based on visual inspection (Note: this assessment could be improved with computer assistance as a future outlook). The prompts are sorted by their distance to the topic cluster (1st to 4th), and whether the samples landed inside or outside the cluster is indicated with (in) and (out).

The table does not reveal a clear trend for consistently well-performing prompt formulations. The sorting of individual samples appears quite random. None of the samples landed within their corresponding topic clusters.

The figures 12, 13, and 14 depict the comparison of the different tested prompts analyzed by style. In table 7, the resulting outcomes are listed based on visual inspection (Outlook: this evaluation could be improved with computer assistance as a future outlook). The prompts are sorted by their distance to the style cluster (1st to 4th), and it is noted whether the samples landed inside or outside the cluster.

Similarly, no clear trend for consistently well-performing prompt formulations can be discerned from the table. The sorting of the individual samples appears quite random. Some of the samples landed within their respective style cluster; however, these are in the minority.

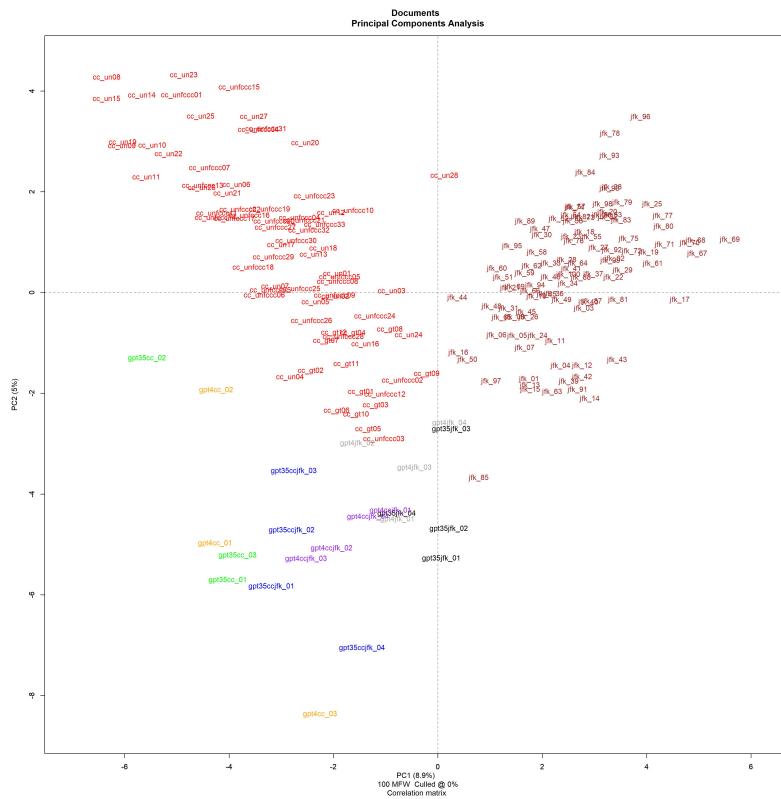


Figure 9: Topic Modelling on cc, jfk and all gpt samples related to jfk

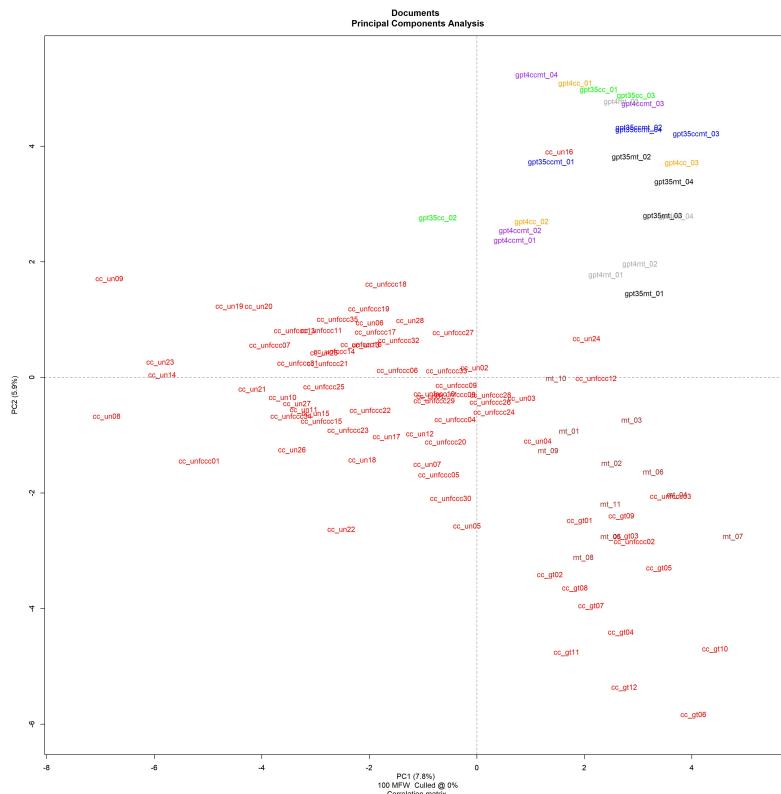


Figure 10: Topic Modelling on cc, mt and all gpt samples related to mt

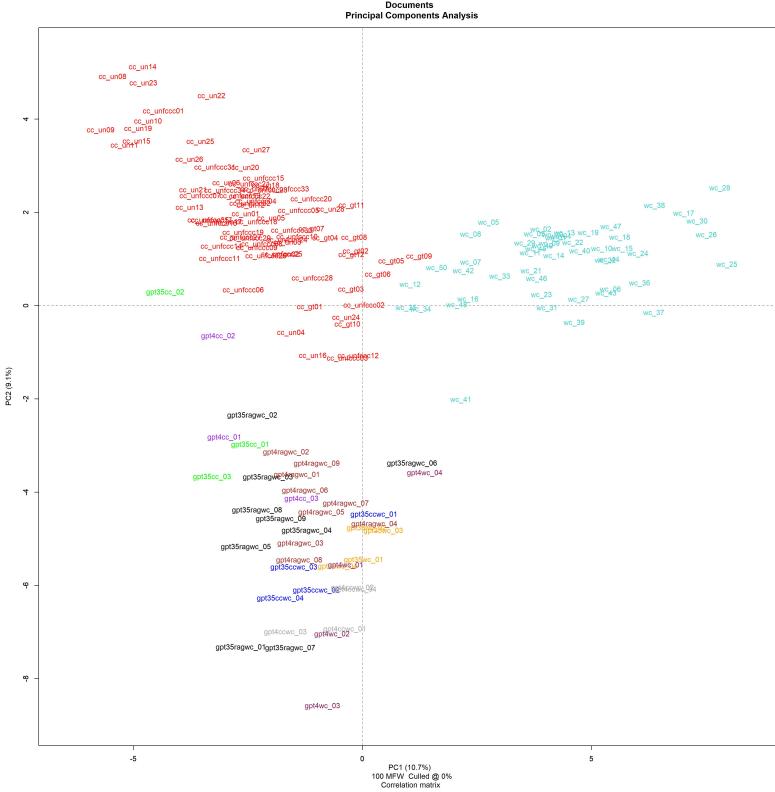


Figure 11: Topic Modelling on cc, wc and all gpt samples related to wc. gpt35ragwc (black) and gpt4ragwc (brown) are prompts augmented with original Speeches from Winston Churchill

Acronym	color	1st	2nd	3rd	4th
gpt35.cc	green	02 (out)	03 (out)	01 (out)	
gpt35.jfk	black	03 (out)	02 (out)	04 (out)	01 (out)
gpt35_mt	black	01 (out)	03 (out)	04 (out)	02 (out)
gpt35_wc	yellow	02 (out)	03 (out)	01 (out)	04 (out)
gpt35_ccjfk	blue	03 (out)	02 (out)	01 (out)	04 (out)
gpt35_ccmt	blue	01 (out)	03 (out)	04 (out)	02 (out)
gpt35_ccwc	blue	01 (out)	03 (out)	02 (out)	04 (out)
gpt4_cc	yellow	02 (out)	01 (out)	03 (out)	
gpt4_jfk	grey	04 (out)	03 (out)	02 (out)	01 (out)
gpt4_mt	grey	01 (out)	02 (out)	04 (out)	03 (out)
gpt4_wc	purple-red	04 (out)	01 (out)	02 (out)	03 (out)
gpt4_ccjfk	purple	01 (out)	04 (out)	02 (out)	03 (out)
gpt4_ccmt	purple	01 (out)	02 (out)	03 (out)	04 (out)
gpt4_ccwc	grey	02 (out)	04 (out)	01 (out)	03 (out)

Table 6: Prompt Comparison Evaluation on Topic by eye. 1st to 4th describes which sample is closest to the topic cluster. (in) indicates that the sample has landed in the topic cluster

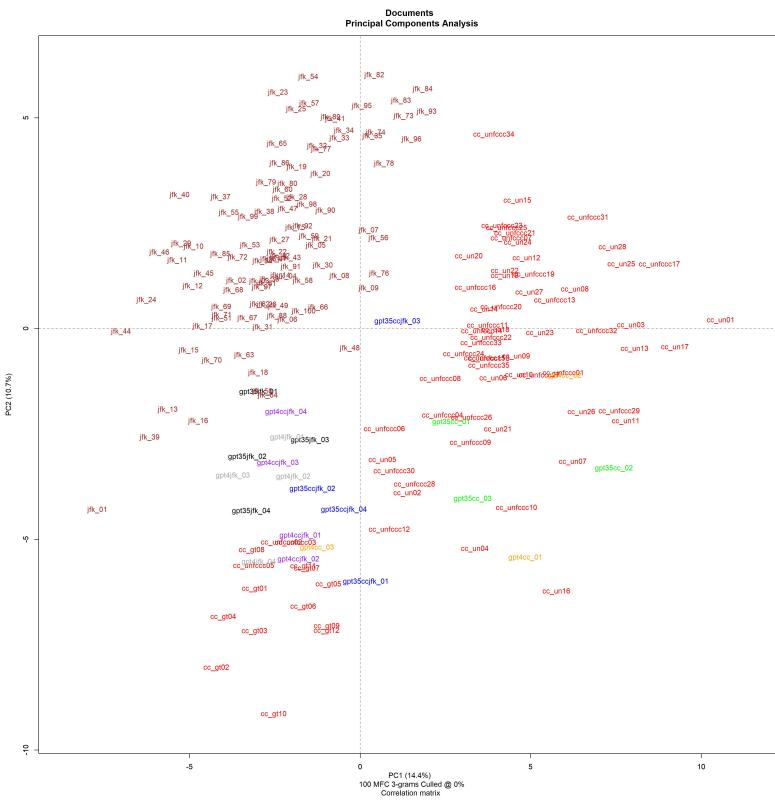


Figure 12: PCA (corr.) on Feature Char 3-Gram on cc, jfk and all gpt samples related to jfk

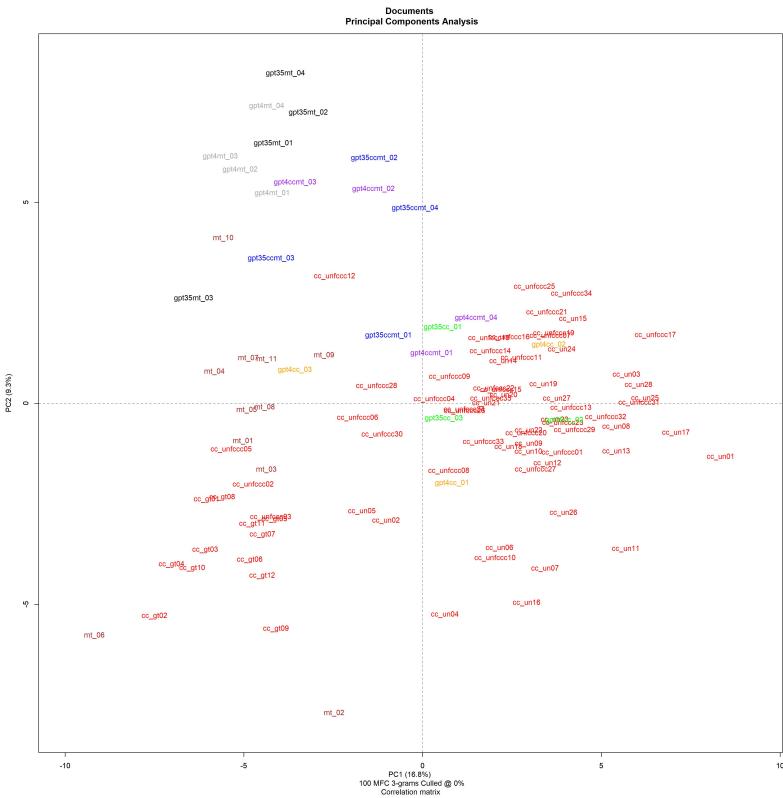


Figure 13: PCA (corr.) on Feature Char 3-Gram on cc, mt and all gpt samples related to mt

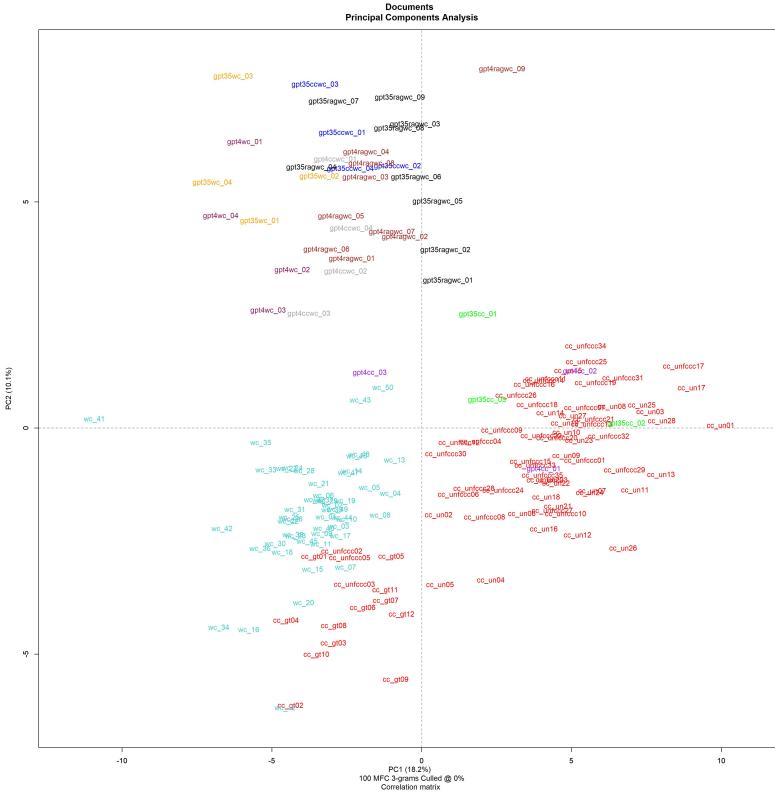


Figure 14: PCA (corr.) on Feature Char 3-Gram on cc, wc and all gpt samples related to wc. gpt35ragwc (black) and gpt4ragwc (brown) are prompts augmented with original Speeches from Winston Churchill

Acronym	color	1st	2nd	3rd	4th
gpt35_cc	green	03 (in)	02 (in)	01 (in)	
gpt35_jfk	black	01 (in)	02 (out)	03 (out)	04 (out)
gpt35_mt	black	03 (in)	01 (out)	02 (out)	04 (out)
gpt35_wc	yellow	01 (out)	04 (out)	02 (out)	03 (out)
gpt35_ccjfk	blue	03 (out)	02 (out)	04 (out)	01 (out)
gpt35_ccmt	blue	03 (in)	01 (out)	04 (out)	02 (out)
gpt35_ccwc	blue	04 (out)	02 (out)	01 (out)	03 (out)
gpt4_cc	yellow	02 (in)	01 (in)	03 (out)	
gpt4_jfk	grey	01 (out)	03 (out)	02 (out)	04 (out)
gpt4_mt	grey	01 (out)	02 (out)	03 (out)	04 (out)
gpt4_wc	purple-red	03 (out)	02 (out)	04 (out)	01 (out)
gpt4_ccjfk	purple	04 (out)	03 (out)	01 (out)	02 (out)
gpt4_ccmt	purple	01 (out)	03 (out)	04 (out)	02 (out)
gpt4_ccwc	grey	03 (out)	02 (out)	04 (out)	01 (out)

Table 7: Prompt Comparison Evaluation on Style by eye. 1st to 4th describes which sample is closest to the style cluster. (in) indicates that the sample has landed in the style cluster

7.4 Compare augmented prompts

In Figures 11 and 14, prompts augmented with original texts from Winston Churchill were tested alongside the simple prompts. In Figure 11, it can be observed that gpt35ragwc (black) and gpt4ragwc (brown) are slightly closer to the cc cluster compared to their counterparts gpt35ccwc (blue) and gpt35ccwc (grey). This can also be confirmed for gpt35ragwc (black) in Figure 11, but not for gpt4ragwc (brown). None of the prompts land in the corresponding clusters.

Overall, the augmented prompts did not stand out as strongly as expected. They were not sufficient to generate speeches indistinguishable from those by ChatGPT.

7.5 Classify ChatGPT speeches between climate change and political figure speeches

In Figures 9, 10, 11, 12, 13, and 14, the various samples generated by ChatGPT were compared with the speeches on climate change and those of historical politicians. Only in a few cases did the ChatGPT samples meet expectations by landing in the corresponding style clusters and thus being indistinguishable from the originals. However, for most samples, this was not the case. On the contrary, the ChatGPT samples in Figures 9, 10, and 11 formed their own clusters, which could be clearly distinguished from the rest of the samples. In Figures 12, 13, and 14, ChatGPT clusters are also evident, although they exhibit slightly greater overlap with the remaining samples, although not as much as expected.

In summary, it can be said that the samples generated by ChatGPT can be distinguished quite well using stylometry from both speeches on climate change and those of historical politicians. This provides a small selected insight into the ability, or rather underdeveloped ability, of ChatGPT to mimic the style of authors and ultimately provides an approach to automatically detect ChatGPT-generated texts.

All data, R scripts used for data processing and analysis, and all results can be viewed in the Git repository (<https://github.com/MaxOmlor/Comparative-Analysis-of-Large-Language-Models-Stylistic-Responses>).

8 Conclusion

Here is a summary of the key findings from this project, organized according to the defined granular questions:

For comparing topics, among the methods investigated, the PCA (corr.) on the Feature Topic Modelling (see figure 6) method is recommended. Using this method, the speeches on climate change could be well distinguished from those of the historical politicians.

For comparing styles, among the methods investigated, PCA (corr.) on the Feature Char 3-Gram (see figure 7) performed the best. This method stood out primarily due to the poor performance of the other methods investigated. Therefore, as an outlook, further examination of the data and exploration of additional style methods, such as the frequency of grammatical constructs, could be considered at this point.

From the examined prompt formulations, none clearly outperformed the others, and thus, no statement can be made about a particularly good formulation. As a future outlook, further prompts could be tested and compared. Additionally, a systematic approach to developing and optimizing prompts could be developed and tested at this stage.

The augmented prompts performed slightly better in some cases than the simple prompts. However, the overall improvement remains very limited and also did not lead to text generations that cannot be distinguished from the original samples.

For contextualizing the samples generated by ChatGPT, it can be noted that in the vast majority of cases, the generated samples were very distinguishable from the original texts using the simple stylometric methods discussed here. In most cases, the generated samples formed their own clusters that were clearly distinguishable from the rest of the samples.

The analyzed data have also raised several questions. What causes the proximity between the subsets mt and gt? Why is there such a large variation within the mt subset? Why do the samples from the gt subset differ so significantly from those of the un and unfccc subsets, and why is there not something similar observed between un and unfccc? Further examination of the individual text samples could shed light on these questions.

Regardless of these open avenues, the results provide a small glimpse into ChatGPT's ability to mimic the style of authors and how stylometric methods can successfully be used to investigate such cases.

References

- [1] Universität Trier. *Computerlinguistik und Digital Humanities*. 2024. URL: <https://www.uni-trier.de/universitaet/fachbereiche-faecher/fachbereich-ii/faecher/computerlinguistik-und-digital-humanities/infos-fuer-studieninteressierte/was-sind-digital-humanities> (visited on 03/14/2024).
- [2] John F. Kennedy Presidential Library Museum. *John F. Kennedy: Speeches*. 2024. URL: <https://www.jfklibrary.org/archives/other-resources/john-f-kennedy-speeches> (visited on 03/10/2024).
- [3] Margaret Thatcher Foundation. *Search*. 2024. URL: <https://www.margaretthatcher.org/search?dt=4&imp=1i&w=keyspeech> (visited on 03/10/2024).
- [4] International Churchill Society. *Speeches*. 2024. URL: <https://winstonchurchill.org/resources/speeches/> (visited on 03/10/2024).
- [5] United Nations. *Speeches on Climate Action*. 2024. URL: <https://www.un.org/en/climatechange/speeches> (visited on 03/10/2024).
- [6] The Environment Show. *Greta Thunberg – the future speaks*. 2020. URL: <https://www.environmentshow.com/greta-thunberg-speeches/> (visited on 03/10/2024).
- [7] United Nations Climate Change participants. *COP 26 Speeches and statements*. 2021. URL: <https://unfccc.int/cop-26/speeches-and-statements#List-of-Speakers-for-the-Resumed-High-Level-Segment> (visited on 03/10/2024).
- [8] Manuel Burghardt. *Introduction to Digital Humanities - Finding Textual Similarities: Stylometry*. Vorlesung gehalten an der Universität Leipzig. 2024.
- [9] Dr. Ing. Andreas Niekler. *Introduction to Digital Humanities - Topic Models*. Vorlesung gehalten an der Universität Leipzig. 2024.
- [10] Song Feng, Ritwik Banerjee, and Yejin Choi. "Syntactic stylometry for deception detection". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2012, pp. 171–175.
- [11] Glory O. Adebayo and Roman V. Yampolskiy. "Estimating Intelligence Quotient Using Stylometry and Machine Learning Techniques: A Review". In: *Big Data Mining and Analytics* 5.3 (2022), pp. 163–191. DOI: 10.26599/BDMA.2022.9020002.
- [12] Aylin Caliskan-Islam et al. "De-anonymizing Programmers via Code Stylometry". In: *24th USENIX Security Symposium (USENIX Security 15)*. Washington, D.C.: USENIX Association, Aug. 2015, pp. 255–270. ISBN: 978-1-939133-11-3. URL: <https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/caliskan-islam>.
- [13] Geri Mileva. *The Ultimate AI Prompt Optimization Guide for 2024*. 2024. URL: <https://guides.temple.edu/stylometryfordh> (visited on 03/14/2024).

9 AI Tool Usage

This section documents the use of AI tools within this project. AI tools were employed for various tasks within different processes throughout this project. Explanations of the tasks and processes will be listed below, along with details of their implementation.

9.1 Code generation for a problem description

AI-supported code generation, exemplified by the development of the function 'encode_sentence_len':

You:

schreibe mir eine funktion in r, die einen text entgegen nimmt und jeden satz durch ein wort ersetzt. dazu soll der text in eine liste von dessen sätzen zerlegt werden. jeder satz soll dann in ein wort codiert werden. das wort soll so aufgebaut sein: f'len{sentence_length}'dersoentstandenetextsollzurückgegebenwerden

Listing 1: Mein R-Code

```
library(stringr)

# Funktion, die einen Text entgegennimmt und jeden Satz durch ein kodiertes Wort ersetzt
encode_sentences <- function(text) {
  # Zerlege den Text in Sätze
  sentences <- unlist(str_split(text, "(?=<\\.)\\s", perl=TRUE))

  # Kodiere jeden Satz durch ein Wort basierend auf seiner Länge
  encoded_sentences <- sapply(sentences, function(sentence) {
    sentence_length <- nchar(sentence)
    return(paste0("len", sentence_length))
  })

  # Kombiniere die kodierten Wörter zu einem Text
  encoded_text <- paste(encoded_sentences, collapse="■")

  return(encoded_text)
}

# Beispieltext
text <- "Dies■ist■ein■Beispieltext.■Er■besteht■aus■mehreren■Sätzen.■Jeder■Satz■wird■"

# Kodiere den Text
encoded_text <- encode_sentences(text)

# Ausgabe des kodierten Texts
print(encoded_text)
```

You:

bitte berücksichtige auch, dass .!? von einem zeilenumbruch gefolgt sein können

Listing 2: Mein R-Code

```
library(stringr)

encode_sentences <- function(text) {
  # Zerlege den Text in Sätze unter Berücksichtigung von .!?
  sentences <- unlist(str_split(text, "(?<=[.!?])([\s]|[\r?\n]+)", perl=TRUE))

  # Kodiere jeden Satz durch ein Wort basierend auf seiner Länge
  encoded_sentences <- sapply(sentences, function(sentence) {
    sentence_length <- nchar(sentence)
    return(paste0("len", sentence_length))
  })

  # Kombiniere die kodierten Wörter zu einem Text
  encoded_text <- paste(encoded_sentences, collapse="■")

  return(encoded_text)
}

# Beispieltext mit Zeilenumbrüchen
text <- "Dies■ist■ein■Beispieltext.■Er■besteht■aus■mehreren■Sätzen.\nJeder■Satz■wird

# Kodiere den Text
encoded_text <- encode_sentences(text)

# Ausgabe des kodierten Texts
print(encoded_text)
```

You:

bitte passe das encoding: # Kodiere jeden Satz durch ein Wort basierend auf seiner Länge
encoded_sentences <- sapply(sentences, function(sentence) sentence_length <- nchar(sentence)
return(paste0("len", sentence_length)))

so an, dass nicht die anzahl der chars gezählt wird, sondern die anzahl der wörter. zerlege dazu zunächst jeden satz in eine liste aus wörtern

etc.

This iterative method was applied to practically the entire resulting code, with some manual interventions and adjustments.

9.2 Error search by specifying code and error.

AI supported error searching in the example of the function 'gen_agg_hists'.

You:

das ist meine aktuelle funktion: [gen_agg_hists code]

warum wird mir kein plot ausgegeben?

ChatGPT:

Es sieht so aus, als ob Ihr Plot nicht angezeigt wird, weil Sie ihn nicht explizit anzeigen. In Ihrer Funktion wird der Plot zwar erstellt und als Datei gespeichert, aber er wird nicht auf dem Bildschirm angezeigt. Um den Plot anzuzeigen, können Sie print(p) oder einfach p am Ende der Funktion verwenden. Hier ist Ihre Funktion mit der Anpassung:

[vermeindlich verbesserter gen_agg_hists code]

Durch das Hinzufügen von print(p) wird der Plot direkt nach seiner Erstellung angezeigt.

etc.

Error searching with ChatGPT is usually only successful in very obvious cases and was therefore dominated by manual search and manual attempts. However, an attempt to involve ChatGPT in the error search was made in a large portion of the errors.

9.3 Feature expansion by providing the previous code passage and a verbal description of a desired feature.

This follows the process of code generation for a problem description, with additional specification of the existing code. It's also an iterative process with manual interventions and adjustments.

This process was used at several stages of feature development.

9.4 Translating text passages.

This process is exemplified by a translated paragraph.

You:

Unabhängig dieser offenen Ansätze bieten die Ergebnisse einen kleinen Einblick in die Fähigkeit von ChatGPT den Stil von Autor*innen imitieren und wie Stylometrische Methoden erfolgreich dazu genutzt werden können solche Fälle untersuchen.

ChatGPT:

Regardless of these open avenues, the results provide a small glimpse into ChatGPT's ability to mimic the style of authors and how stylometric methods can successfully be used to investigate such cases.

These translations could be adopted with minimal modifications and were used for almost all paragraphs in this report.

9.5 Generating a paragraph

This process is illustrated using the example of the introduction.

You:

gegeben dieser research question: How does an advanced language model like ChatGPT respond to requests for generating speeches on modern topics, such as climate change, in the style of historical figures like Winston Churchill, Margaret Thatcher, and John F. Kennedy, and can stylometric analyses distinguish the differences between these AI-generated texts and authentic speeches of these figures, as well as real speeches on the topic of climate change?

und dieser conclusion: Zum Vergleich der Themen bietet sich unter den untersuchetne Methoden, die Methode PCA (corr.) auf dem Feature Topic Modelling (vgl. figure 6) an. Mit dieser Methode konnten die Klimawandel reden gut von denen der historischen Politiker*innen unterschieden werden. Fur den Vergleich der Stile hat unter den untersuchten Methoden die PCA (corr.) auf dem Feature Char 3-Gram (vgl. figure 7) am besten abgeschnitten. Diese Methode setzte sich vor allem wegen der schlechten Performance der anderen untersuchten Methoden durch. Entsprechend konnten an dieser Stelle als Ausblick die Daten noch einmal genauer untersucht und weitere Stil Methoden betrachtet werden. Aus den untersuchten Prompt Formulierungen hat sich keine deutlich von den anderen abgesetzt und somit kann keine Aussage über besonders gute Formulierung gemacht werden. Hier konnten als Ausblick weitere Prompts getestet und verglichen werden. Zusätzlich konnte an dieser Stelle ein systematischer Ansatz zur Entwicklung und Optimisierung der Prompts entwickelt und getestet werden. Die augmentierten Propmts konnten in einigen Fällen etwas besser abschneiden, als die einfachen Prompts. Die Verbesserung bleibt aber insgesamt sehr überschaubar und hat ebenfalls nicht zu Text Generierungen geführt, die sich von den Orginal Samples nicht unterscheiden lassen. Zur Einordnung der von ChatGPT generierten Samples lässt sich feststellen, dass sich die generierten Samples in den aller meisten Fällen sehr gut mit den einfachen hier diskutierten Stylometrischen Methoden von den Orginaltexten unterscheiden ließ. In den Meisten Fällen bildeten die Generierten Samples eigene Cluster die sich klar von den restlichen Samples abgrenzen ließen. Durch die untersuchten Daten wurden darüber hinaus noch einige Fragen aufgeworfen. Wodurch entsteht die Nähe zwischen den Subsets mt und gt? Warum besteht innerhalb von mt eine so große Streuung? Warum setzen sich die Samples das Subsets gt von denen der Subsets un und unfccc so deutlich ab, bzw. warum ist nicht etwas vergleichbares zwischen un und unfccc zu beobachten? Hier konnten die einzelnen Text Samples noch einmal genauer untersucht werden. Unabhängig dieser offenen Ansätze bieten die Ergebnisse einen kleinen Einblick in die Fähigkeit von ChatGPT den Stil von Autor*innen imitieren und wie Stylometrische Methoden erfolgreich dazu genutzt werden können solche Fälle untersuchen.

schreibe mir eine Introduction für eine wissenschaftliche arbeit

ChatGPT:

Die fortschreitende Entwicklung von Sprachmodellen auf Basis künstlicher Intelligenz hat in den letzten Jahren zu bemerkenswerten Fortschritten geführt und eröffnet neue Möglichkeiten für die Generierung von Texten. Insbesondere fortschrittliche Sprachmodelle wie ChatGPT haben das Potenzial, auf komplexe Anfragen zu reagieren und hochwertige Texte zu erstellen, die dem Stil und der Thematik historischer Persönlichkeiten entsprechen.

In dieser Studie untersuchen wir, wie sich ein fortgeschrittenes Sprachmodell wie ChatGPT auf Anfragen zur Generierung von Reden zu modernen Themen wie dem Klimawandel verhält. Dabei liegt unser Fokus darauf, ob ChatGPT in der Lage ist, Reden im Stil bekannter historischer Figuren wie Winston Churchill, Margaret Thatcher und John F. Kennedy zu produzieren. Zudem analysieren wir, ob stilometrische Methoden die Unterschiede zwischen den von ChatGPT generierten Texten und authentischen Reden dieser Persönlichkeiten sowie realen Reden zum Thema Klimawandel erkennen können.

Die Erkenntnisse aus dieser Untersuchung sind von hoher Relevanz, da sie nicht nur Einblicke in die Leistungsfähigkeit von Sprachmodellen wie ChatGPT bieten, sondern auch die Anwendung stilometrischer Analysen zur Bewertung von KI-generierten Texten beleuchten. Darüber hinaus können die Ergebnisse dazu beitragen, die Weiterentwicklung und Anwendung von Sprachmodellen in verschiedenen Bereichen zu unterstützen und die Diskussion über den Einfluss von KI auf die Erstellung von Texten zu vertiefen.

Diese Studie ist in mehrere Abschnitte unterteilt. Zunächst geben wir einen Überblick über den aktuellen Stand der Forschung auf diesem Gebiet. Anschließend beschreiben wir unsere Methodik und die verwendeten Datenquellen. Danach präsentieren wir die Ergebnisse unserer Analyse und diskutieren deren Implikationen. Schließlich ziehen wir Schlussfolgerungen und geben einen Ausblick auf mögliche zukünftige Forschungsrichtungen.

This process was exclusively applied to create the introduction. The generated introduction was largely adopted, with some additional content added in certain parts.

9.6 Summaries of texts or web pages

This process is illustrated using the overview of DH as an example.

You:

lese diese Webseite: <https://www.uni-trier.de/universitaet/fachbereiche-faecher/fachbereich-ii/faecher/computerlinguistik-und-digital-humanities/infos-fuer-studieninteressierte/was-sind-digital-humanities>

und beantworte mir diese Fragen: Paint the big picture of DH (what is it, how is it different from traditional humanities, etc. – use references here!)

ChatGPT:

Digital Humanities (DH) vereinen Geisteswissenschaften mit Informatik, fokussierend nicht nur auf Philologien sondern auf alle Geistes- und Kulturwissenschaften. DH nutzen digitale

Technologien für Digitalisierung, Archivierung, Repräsentation, Visualisierung und Analyse geisteswissenschaftlicher Daten. Während traditionelle Geisteswissenschaften oft manuelle Analysen durchführen, ermöglichen DH automatisierte Verfahren und tiefere Einblicke durch die Verarbeitung großer Datenmengen

This process was exclusively used to generate this overview. The overview from ChatGPT was adopted with only minimal adjustments.

9.7 Structural recommendations, such as the organization of a chapter

This process is exemplified by potential challenges in using stylometry and topic modeling.

You:

was könnten schwierigkeiten dieser methoden:

Stylometry

Stylometry wird im Allgemeinen dazu verwendet Ähnlichkeiten in Texten zu identifizieren. Dafür können unterschiedliche quantifizierbare textuelle Features analysiert werden, wie die im Text enthaltenen Wörter, die Länge der Sätze, Frequencies von Buchstaben(gruppen) und Wort(gruppen), Frequencies von Grammatikalischen Konstruktionen, Frequencies von seltenen Wörtern und viele weitere. Daraus lassen sich Feature Vektoren ableiten, die einen gegebenen Text über dessen Features beschreiben. Diese Vektoren können dann über verschiedene Distanz Maße oder durch die Repräsentation mehrerer Vektoren in einem PCA analysiert und verglichen werden, um z.B. eine Aussage über die Stilistische Ähnlichkeit der gegebenen Texte zu treffen.

In diesem Projekt wurden verschiedene Features, wie Bag of Words, 3-Gramme über Buchstaben, die Länge der Sätze, die Punktuation und einige Features die den Topic modelling zugeordnet (und auf die dort nochmal eingegangen wird) werden können, verwendet.

Bag of Words beschreibt ein Verfahren, welches die vorkommen jedes einzelnen wortes in einem gegebenen Textes zählt und diese Anzahlen in einem Vektor zusammenfasst. 3-Gramme über Buchstaben zählen statt ganzer wörter die vorkommen von 3 aufeinander folgenden Buchstaben. Diese Anzahlen werden ebenfalls in einem Vektor zusammengefasst.

Topic Modelling

Topic Modelling wird benötigt, um die Semantische Bedeutung von Wörter zu bestimmen. Da es bspw. Wörter gibt, die je nach Kontext mehrere Bedeutungen haben können, reicht eine einfache Frequency Analyse nicht aus um eine Aussage über die Semantische Beteutung treffen zu können. Zu diesem zweck werden Topic Models eingesetzt. diese modellieren Topics als listen von Wörtern, welche in einem gegebenen Text Corpus statistisch auffällig oft zusammen auftreten.

In diesem Projekt wurden Topic Models verwendet um daraus Features für eine Frequency Analyse zu erstellen.

im bezug auf diese Research question: How does the advanced language model ChatGPT respond to requests for generating speeches on a modern topic like climate change, in the style of historical figures like Winston Churchill, Margaret Thatcher, and John F. Kennedy, and can stylometric analyses distinguish the differences between these AI-generated texts and authentic speeches of these figures, as well as real speeches on the topic of climate change? sein?

ChatGPT:

Einige mögliche Schwierigkeiten bei der Verwendung von Stylometrie und Topic Modelling in Bezug auf die gegebene Forschungsfrage könnten sein:

Repräsentativität der Daten: Es könnte schwierig sein, einen ausreichend großen und repräsentativen Datensatz historischer Reden von Winston Churchill, Margaret Thatcher und John F. Kennedy zu finden, um sie mit den AI-generierten Texten und echten Reden zum Thema Klimawandel zu vergleichen. Eine unzureichende Datenmenge könnte die Zuverlässigkeit der Stylometrie und des Topic Modelling beeinträchtigen.

Komplexität der Stilanalyse: Die Stilanalyse historischer Reden und die Unterscheidung zwischen verschiedenen Stilen könnten komplex sein, insbesondere wenn es darum geht, subtile stilistische Unterschiede zwischen den Reden der verschiedenen historischen Figuren zu identifizieren. Dies erfordert möglicherweise fortschrittliche Stylometrie-Techniken und eine detaillierte Untersuchung der verschiedenen stilistischen Merkmale.

Interpretation der Topic-Modelle: Die Interpretation der Topic-Modelle kann herausfordernd sein, da sie oft abstrakte Konzepte repräsentieren und eine manuelle Überprüfung erfordern können, um die Bedeutung der identifizierten Themen zu verstehen. Darüber hinaus kann die Erstellung von Features aus den Topic-Modellen und deren Integration in die Stilanalyse zusätzliche Schwierigkeiten mit sich bringen.

Vergleichbarkeit der Ergebnisse: Es kann schwierig sein, die Ergebnisse der Stylometrie und des Topic Modelling zwischen den verschiedenen Arten von Texten (AI-generierte Texte, historische Reden, echte Reden zum Thema Klimawandel) zu vergleichen und Schlussfolgerungen über die Unterschiede in Stil und Inhalt zu ziehen. Eine sorgfältige Validierung und Interpretation der Ergebnisse ist erforderlich, um aussagekräftige Schlussfolgerungen zu ziehen.

Berücksichtigung von Kontext und Intention: Bei der Analyse der Stilunterschiede zwischen den AI-generierten Texten und den authentischen Reden ist es wichtig, den Kontext und die beabsichtigte Wirkung der Reden zu berücksichtigen. Dies könnte eine zusätzliche Herausforderung darstellen, da der Kontext historischer Reden und die Absichten ihrer Verfasser möglicherweise nicht vollständig bekannt oder verstanden sind.

Indem diese Schwierigkeiten adressiert und sorgfältig angegangen werden, kann eine fundierte Analyse durchgeführt werden, um die Forschungsfrage zu beantworten und Erkenntnisse über die Fähigkeiten von ChatGPT sowie die Unterschiede zwischen AI-generierten und authentischen Reden zu gewinnen.

This response inspired the corresponding paragraph in Section 4.

This process has also influenced this report by suggesting the use of 3-gram chars for structural analysis.