Critical Discourse Studies (CDS) of Social Media: Taking a Greta Thunberg's Speech on YouTube as a Case Study

Seminar "Critical Social Median Analysis", Winter Term, 2020/21

Zheng Yao
Yuxin Liu
Freie Universität Berlin
Berlin, Germany
zheng.yao@fu-berlin.de
yuxin.liu@fu-berlin.de

ABSTRACT

Nowadays Social media has become an important information platform in our society. People use social media to connect to eachother, express their opinions, share their news, and discuss hot issues of society. With the development of social media, The influence on society and people's daily life has become stronger. This research takes a speech of a Swedish environmental activist Greta Thunberg on YouTube as an example. Taking critical discourse analysis as a theoretical perspective. With qualitative analysis, we found the attitudes and intention behind the video. At the same time, by analyzing 27,324 online comments of the video, we found that the netizens have different attitudes toward Greta and environmental protection. With machine learning methode and Quantitative analysis, and using descriptive statistical methods. The authors believe that social media is playing an increasingly important role in expressing public opinions. This research helps us understand the attitudes of Internet users on the topic such as environmental protection.

KEYWORDS

CDS, Social Media, Greta Thunberg, machine learning, clustering, language model

CONTENTS

1 INTRODUCTION

Social media has developed tremendously in the Internet 2.0 era. Various online social media, such as Facebook, Twitter, Instagram, YouTube in the United States, and various domestic online forums, blogs, Sina Weibo, Tencent WeChat in China, etc. have emerged one after another. Regarding social media, Boyd and Ellison(2008: p. 210-230)[7] defined it as "a service based on the Internet, which allows users to create a public or semi-public personal profile in the system. The shared relationship is used to generate a user list, and the related information of other user members associated with it can be viewed in this system. "User-generated content, shared information and its interactive communication characteristics have led to the rapid development of social media and gradually infiltrating public life. In all aspects, several major international and domestic social media, such as Facebook, Twitter, Sina Weibo, WeChat, etc., have hundreds of millions of users, which to a certain extent means that users who use social media are connected to each other in some form As Hanna (2011: p. 265)[11] and other scholars said, "We are all connected through social media."

Online social media is playing an increasingly important role in people's daily interpersonal communication, and it also has an important impact on environmental protection. The topic of environmental protection is actively discussed on social networks, and there are many different views and opinions. This article uses Greta's environmental protection speech as an example to conduct a critical discourse analysis of the online comments she generates to reveal the perspectives and emotional attitudes reflected in the discourse.

The speeches of Greta and the discussion about her in personal have always been the hot spot on the topic of environmental protection. Since August 2018, This Swedish environmental activist has become the leader of the global youth anti-climate change movement in just a few months. She has earned a series of honors and awards meanwhile also attracted a lot of controversy and

criticism, such as "climate protection paranoia", "climate activists using children for political propaganda", and "questioning some of her behavior Not environmentally friendly"[5]. Therefore, studying Greta's speech videos and comments will help to understand the public's attitude towards environmental protection.

2   RESEARCH DESIGN

2.1   Motivation and Research Goals

Greta Tintin Eleonora Ernman Thunberg is a Swedish environmental activist and internationally known for challenging world leaders to take immediate action against climate change. On 20 Aug. 2018, Thunberg started to skip school to protest outside Swedish parliament for more actions against climate change. From then on, her idea and motivation on urging world leaders to make the climate a priority have been propagated through speeches and interviews by mainstream media. Greta Thunberg's speech at the U.N. Climate Action Summit[9] targeting world leaders attracted high attention, so that this speech video has the most views among all speeches by Thunberg on the global notable news platform including the Guardian News, PBS News Hour etc. This project focuses on the speech itself and the response from the video audiences over time. Through the Critical Discourse Analysis (CDA) (Fairclough 1995)[10] on the speech transcript, the reason for the heated debate caused by this video will be investigated. With the help of the Appraisal Theory (Martin and White, 2005)[15], the attitudes expressed in the comments will be studied. The aim of this research is to investigate:

- Do speeches by Greta Thunberg really raise people's concern for climate change, or be counterproductive?
- What is the attitude of the audiences towards her speeches: agree, disagree or ambiguous?
- What are the hit issues regarding the certain speeches by Thunberg?

2.2   Research Questions

Based on the research goals, the research question with consideration on the available materials and techniques have been designed as following:

- How many comments are supportive or opposite towards Thunberg?
- How many comments are supportive or opposite towards environment protection?
- Any overlap among these two groups?

Of the four research questions mentioned above, the first and the second question target at linguistic and emotional expressions reflected in the comments based on Appraisal Theory (Martin and White, 2005)[15]. The focus of these two questions is slightly different. The first question focuses on the attitude towards the speaker Greta Thunberg, while the second question pays more attention to the comments centered on the environment protection. The obtained result will be connected to the critical discourse analysis on the speech text, where the discursive strategies employed by the speaker for the specific event context and the counterpart of the speech, as well as the social behavior it maps, will be analyzed.

The third question is a further research based on the previous studies, where the correlation between the results from two research questions above will be investigated qualitatively.

2.3   Data Collection

2.3.1 Choose the Video. A satisfactory video for this study should fulfil the following three criteria. Firstly, the chosen video should be a famous speech by Greta Thunberg which contains controversial and discussable materials for potential intense discussion. Secondly, the video should be published in the early phase of Thunberg's road on environment protection, so that the comments before and after each major event by Greta Thunberg can be included for comparative study. Lastly but not least, the chosen video should have provoked heated debates amongst audiences, reflected in the number of comments, it should be the videos with more than 5000 reviews, and with at least 10% of the comments replied by another user. Therefore, Thunberg's speeches targeting world leaders which were reported by the Britain Guardian Media Group were chosen. The chosen videos have the largest views and fulfill the three requirements above. The comment data was obtained by the digital tool YouTube Data Tools and this tool will be discussed in detail.

2.3.2 Digital Methods and Tools. YouTube Data Tools (YTDT)[3] was used to scrawl the desired data from YouTube platform. This digital tool has been applied in a study by Hirsbrunner on how YouTube users understand, negotiate and appropriate science-related knowledge on YouTube. It has been proven to be valid for argument analysis (Hirsbrunner 2021)[12]. YTDT combines the YouTube API and embedded sort and filter functions, enabling the search with focus on a certain keyword or the search for the video with most views. In this project, the scrawler and the sort function were used to obtain the data for the YouTube video with highest views centered on "Greta Thunberg" in the channel "Guardian". The comment time, comment content, the number of likes and the number of replies for each comment were cached as an excel file for further study.

2.4   Data Processing and Analysis

We use the pipeline provided by our Seminar[2] to process our data. And then use the results for quantitative analysis. First, we input the raw data generated by the YouTube Data API[1] into the first language model and obtain the tensor of comments. Then use the clustering model to cluster the reviews and visualize the results. Finally, we use the results for descriptive statistical analysis. The statistical distribution of the experimental results can be used to analyze users' attitudes towards Greta's speech and attitudes towards environmental protection. We may even get how these atitudes change over time. Based on the analysis and discussion of specific examples, we can summarize the results of data processing statistically, and obtain the proportion distribution of these clusters.

3   PRELIMINARY INSIGHTS

Our research is mainly based on qualitative analysis and supplemented by quantitative descriptive statistics. First, from a theoretical perspective, critical discourse analysis(CDA) is used to analyze the discourse of Greta Thunberg's envirimental speech on YouTube as well as its comments. Secondly, with the help of descriptive statistical methods, the types of online comments and opinions of

YouTube users are statistically classified to analyze the altitude conflicts between them. Finally, we combine with the data obtained from qualitative and quantitative analysis, summarize the discourse characteristics of social media in this case.

We will use the three-dimensional analysis framework of Fairclough (1995: p. 96 102) [10] for critical discourse analysis in order to discover discourse strategies in speeches. In terms of online comments, after applying the pipeline to perform clustering analysis, we will analyze its text style and sentimental attitudes and analyze each of the online reviews based on the relevant framework of Martin and White (2005) [15]'s Appraisal Theory. And analyze these Clusters for classification. Finally, apply descriptive statistical analysis (Mann, 1995)[14] to statistically analyze the use of language and emotional characteristics in online comments, so as to derive the attitudes and trends of online social media regarding environmental protection.

Up to now, we have completed the cluster analysis of the data. Other qualitative work will be completed in the next steps.

As mentioned above, the three-dimensional framework of critical discourse analysis of Fairclough (1995)[10] is the main theoretical analysis framework of this research. According to Fairclough (1995)[10], critical discourse analysis can have three stages, namely Description, Interpretation and Explanation. Specifically, description is the process of text analysis. This is a relatively micro analysis stage that focuses on the text itself, such as style, grammar, syntax, vocabulary, rhetoric, tone, voice, and subject description Micro-analysis of language structure and characteristics; Interpretation is the process of processing analysis. This stage involves text and discourse practice, that is, exploring how discourse is in a specific social environment. How is the next generation, and how it is disseminated and consumed; Explanation has risen to the macro stage of social analysis, this stage includes the participation of Discourse practice and Social-cultural practice , To further reveal social factors, such as politics, economy, culture, religious beliefs, and interest groups, shaping the discourse, and the reproduction and strengthening of the social structure of discourse. From text to discourse to the level of social culture, this is the basic analysis logic of Fairclough's three-dimensional analysis framework.

In addition, this research also draws on the Appraisal Theory framework of Martin and White (2005)[15]. This theoretical framework is used to analyze the emotions and attitudes of online reviews. Appraisal Theory divides language evaluation resources into three categories, namely, attitude, engagement and graduation three subsystems (quoted from Yang, 2010: p. 41)[20]. In the attitude subsystem, according to White (2015: p. 2 3)[18], the theory's assessment of attitudes defines three broad categories, namely, Appreciation and Judgment. Appreciation refers to "evaluating artifacts, entities, events, and states of affairs by referring to aesthetics and other social evaluation systems"; Judgment refers to "assessing human behavior and/or character with reference to morals and other social norms"; Affect refers to Think of "evaluation as an emotional response." Combining with the purpose of this research, the Affect classification of Appraisal Theory is used to analyze the emotion and attitude expression of online reviews, in order to summarize the sentiment and attitude tendencies toward environmental protection in online comments in the research.

4 THEORY AND METHODOLOGY

4.1 3D Critical Discourse Analysis Theory

Critical Discourse Analysis, in short CDA was developed by Fairclough in 1995. It is an approach to analyze any discourse whether in written or spoken form from the perspective of discursive practices in a society. Well-known applicable field are e.g. domination, inequality of basic rights and reveal ideology according to van Dijk (1998a)[6].

The framework of CDA was introduced by Fairclough in 1995[10], which consists of three dimensions, i.e. Text, discourse practice and social practice. In the first stage, the work centers on micro-analysis of linguistic structures and features such as lexical choices, cohesion, coherence and different stylistic features. The analysis of grammar, vocabulary, syntax, rhetoric, inflection etc. are also involved in this stage. In the second dimension, the relation between discourse and social context is of interest. More specifically, how discourse is produced in a particular social context, and how it is transmitted and consumed by the YouTube audiences. The third dimension mainly focuses on the socio-cultural practice which further reveals the role of social factors in shaping the discourse, e.g. political, economic, cultural. Three dimensions are also termed as description, interpretation and explanation in several literatures (Fairclough, 1995)[10]. By adopting this CDA framework, the text is inspected by the social theory of language functioning of ideological and political processes. Several publications have shown the
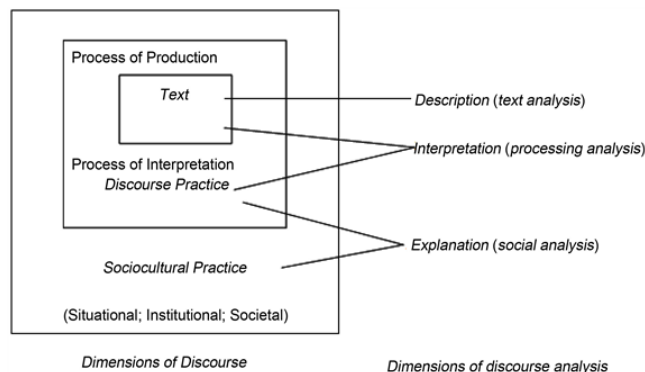

Figure 1: Three-dimensional framework of CDA

applicability of Fairclough's CDA framework to study the speech text which calls for the awareness of the public. Chilton (1996)[19] pointed out that King Martin Luther's speech reflect his ideology in that socio-political background. Sipra et al. further applied this 3D Model to the speech of Martin Luther King 'When I Have Dream' to investigate the hegemonic attitude and discursive practices of the white against the black (Sipra, 2013[17]).

In our work, the 3D Critical Discourse Analysis Theory was applied to analyze the specific structure of speech transcript, aiming to relates the text with broader socio-cultural perspective of discourse and reveal the uneven ideology in the back scene of Thunberg's speech.

4.2 Appraisal Theory

The Appraisal Theory was introduced by Martin and White in 2005[18]. This theoretical framework has been used for sentiment analysis and attitude analysis of the online comments. Zhang et al. have proven its capability to analyze the comments of the YouTube video (Zhang, 2018)[21]. According to the Appraisal Theory, the evaluation resources of the language can be divided into three categories, i.e. attitude, engagement and graduation. White et al. (White, 2015[18]) further defined three classes in the subsection of attitude, namely appreciation, judgment and affect. The subclass "affect" refers to the assessment as an emotional response, which accord with the research goal. Thus, in this work, the subclass "affect" from the Appraisal Theory was used to analyze the sentiment and attitudes revealed by the online comments and to investigate the audiences' attitude toward Greta Thunberg and environmental protection.

5 QUALITATIVE ANALYSIS

The climate change speech of Greta Thunberg at the 2019 UN Climate Action Summit on 23 September 2019 has been analyzed and interpreted by applying Critical Discourse Analysis (CDA). This speech targets at UN and world leaders for their empty words and lack of action on the climate crisis. The analyzed transcript is provided by United Nation (Website: npr). It contains 497 words and 15 paragraphs.

5.1 Text Analysis

The first stage, text analysis, centers on the language analysis. Inspired by Sipra, the analysis was divided into two sections. The first section is linked to the clauses or sentences which highlighted the informative structure (Theme, Rheme, Given and New Information), while the second part focuses on the relationship between the objects of study in the first section. four types of thematic progression

With the help of this progressive study methodology, the features revealed by each single words or clauses enable a further study of the cohesion and coherence of the text as a whole.

Repetition is an obvious feature of Thunberg's speech. The rhetorical question "How dare you!" appears four times in the whole text and respectively once in the first three paragraphs in corresponding to the title of the speech.

- "This is all wrong. I shouldn't be up here. I should be back in school on the other side of the ocean. Yet you all come to us young people for hope. How dare you!"
- "People are suffering. People are dying. Entire ecosystems are collapsing. We are in the beginning of a mass extinction, and all you can talk about is money and fairy tales of eternal economic growth. How dare you!"
- "For more than 30 years, the science has been crystal clear. How dare you continue to look away and come here saying that you're doing enough, when the politics and solutions needed are still nowhere in sight."

Emphasis is often considered as a major function of repetition (Leech 1969)[13]. By repeating the sentence "How dare you?" after each detailed description of world leaders' action against climate change, Thunberg expressed herself firmly and clearly that in her

point of view the world leaders should take responsibility to halt climate change, prioritize the climate saving issues and propose more practical and political solutions for the climate crisis.

The choice on personal pronouns reveals the relationship between the speaker and the listener (Brown and Gilman, 1960)[8]. One evident feature of this speech is the use of second person and the plural form of the first person. By addressing "you" shows solidarity and equality (Brown and Gilman, 1960) and "we" to highlight the speaker as a group instead of a single person (Proctor I-Wen Su 2011:2). Among 38 sentences in Thunberg's speech, "you" appears 28 times to draw the distance between the speaker and audiences (here the UN and world leaders), and also to link up the audience and the listed evidence tightly. For example, the sentences "The eyes of all future generations are upon you. And if you choose to fail us, I say: We will never forgive you." indicates that a good quantity of young generations will not forgive the world leaders and the UN if they fail the expectation of the young people.

The speech as a whole is exquisitely knitted and well organized. as it adopts appropriate words and focusing on the main idea.

5.2 Analyzing Context

The speech of Greta Thunberg carries a complete contextual background. This speech was given at the 2019 UN Climate Action Summit. Greta Thunberg, as a sixteen-year-old environmental activist delivered this speech as a response to a question about the message she has for world leaders. She called for the awareness and focus on the climate change and urged the UN and world leaders to take effective measurements to slow it down. Since according to her experience, they have not done enough for the environmental crisis but tend to leave this problem to the young generations. At the same time, she also criticized their irresponsible behavior and grimness ignoring the voice from the young people but denoting they have done enough.

In Thunberg's speech, UN and world leaders were shaped into an irresponsible image and she was trying to arouse their consciousness on the climate protection.

5.3 Analysis of Discourse Practices and Socio-Political Aspect of Speech

The last section of text analysis is the inspection on the Discourse Practices and Socio-Political Aspect of Speech which is considered as an essential part of the Fairclough's model (Fairclough, 1992)[10]. According to Spira et al., discourse is socially constituted. The importance of socio-cultural and political aspects is providing a complete circumstance which enable the discourses construction (Sipra, 2013)[17].

By using pronoun "We" and "You", Thunberg highlighted the relation among the speaker and audience, setting the world leaders at the opposite site of the young generation. She, as a representative of the young people, expressed their fear of the world with abnormal climate and their anger and pressure resulted from the indifferent attitude and deprivation toward climate change.

5.4 Comments Analysis of YouTube Video

In the last section, the speech transcript was studied with the 3D Critical Discourse Analysis Theory. In this section, the comments

Table 1: repeating the words or sentences from the speech

Comments	Description
Me: Pay her 15% less for her Over acting... Greta Grandma : How dare you!!	Reuse and misinterpretation
Greta: "you have stolen my dreams from me" Starving kids in Africa: "how dare you"	Reuse and misinterpretation
Greta's mum: „Greta, you should been in school" Greta: „How Dare You?"	Reuse and misinterpretation
Greta: *speaking seriously* How dare you have stolen my childhood. We must protect the earth. Other people: ey wanna turn this into a meme?	Reuse and misinterpretation
hOw DaRe yOu!	Reuse and misinterpretation

Table 2: emoji and informal language

Comments	Description
emoji emoji emoji	Pure emoji
you would be evil!, and that I refuse to believe" did u hear that weird laugh in the audience emoji emoji emoji	Emoji combined with text
Howwwwww dareeeeee yeeeeeeee	Informal language
emoji LOL emoji	Informal language with emoji

Table 3: linguistic expression

Comments	Description
I support u greta	Positive and supportive
I think she's brave	Positive
I stAnd with Gretta!!!	Positive and supportive
Oh Mann wie dumm	negative
EP is important, but you are lying	negative but supportive

many of them shows ambiguous attitude, some affect analysis was listed in table 3.

As we learned from the Appraisal Analysis of the comments to her speech, audiences show distinct attitude toward the speaker and the envoronmental protection. 5 classifications of comments were observed according to which attitude was expressed in this comment. They are respectively supportive for Greta and supportive for environmental protection, supportive for Greta but disagree with environmental protection, disagree with Greta but supportive for environmental protection, disagree with Greta and disagree with envoronmental protection, remain nutrual or not expressed. The comments on Greta Thunberg has more quantity than the comments on the environmental protection. Also, the attitude towards Greta Thunberg is highly polarized while the the majority of comments on envoronmental protection shows supportive opinion. In the next section, the sentiment analysis was investigated qualitatively.

After further investigating the replies to the comments, the third feature arise. The interaction among audiences is concentrated in several comments with most of "likes" and disputes are rarely observed. The further study on the replies also confirms this feature. The replies to the comments show high coherence with the replied comment.

6 EXPERIMENT AND RESULT

This section will analyze the video comments. We firstly use the pipeline mentioned in Section 2.4 to cluster the video comments. Then we use descriptive statistical methods to quantitatively analyze the results.

6.1 Linguistic features of comments

We conducted cluster analysis on the collected 27324 online comment corpus data and found that there are three obvious language features overall:

- The use of colloquial expressions and emoticons is frequent.
- The sentence structure of the comments is diverse, which can be short as a few words, and long as a paragraph.
- There are a lot of dirty talks, which are often used for extreme emotion Catharsis.

Let's discuss the three features. Firstly, the frequent use of colloquial expressions and emoticons. Informal English and Internet language are commonly used English styles for online comments

below the chosen speech were analyzed and discussed with the help of the affect analysis from Appraisal Theory. By the 31st December 2020, there were 27324 comments, among which almost were English comments and 198 German comments were recorded.

Three features were observed from them. Firstly, a good quantity of comments was repeating the words or sentences from the speech. Most commonly repeated unit is "How dare you?", which is also the core message she emphasized during the whole speech. However, the context of the comments and the clauses after "How dare you?" mostly differ from her original meaning. As shown in the table 1, more than 90% comments used "How dare you?" as a response in an imaginary conversation to show their disagreement or dislike to Thunberg's acting or her personality. These comments do not concern about the climate change, but only on the speaker itself. This phenomenon is commonly observed in the online comments.

Another feature of the comments is widely used emoji and informal language. As shown in table 2, audiences enjoyed commenting with more than one emoji to expressing or enhancing their emotion. According to the Appraisal Theory by Martin and White, the linguistic expression which reflects intense emotions, is linked to the available resources used for self-expression and reaction in the given context. For example, happy, sad, positive, negative, annoyed and etc. Belong to the system of "affect". Among all the comments,

by YouTube Internet users. At the same time, emoticons are often used to enhance or visualize the emotional attitude of netizens when they make comments. In addition, English grammar is not important, as long as it does not affect basic meaning expression and information exchange. For example, the emoticon "lol" means "laugh/laugh/ridicule", and the emoticons "XD" and ":D" mean "smile/laugh".

Secondly, the diversity of sentence structure and content of online comments. A short comment can be a word that expresses a certain attitude or emotion, while a long comment can consist of one or several paragraphs, which are used to describe one or more opinions of the commenter. The grammar of this type of comment is often more standardized. Most sentence patterns are a combination of long and short sentences, often expressing specific emotions and opinions.

Finally, the third feature is that dirty words are used more often and are often used for extreme emotional expression. In most cases, such offensive language expressions often occur in online quarrels or fierce conflicts of opinions. According to Martin and White (2005: p. 35)[15] in the Appraisal Theory (Appraisal Theory), such language expressions that reflect intense emotions can be classified as the "Affect" subsystem of language, which refers to The ones are resources used to express emotional states and reactions, such as happiness, disgust, sadness, positive, happy, etc. Since Greta was very excited in the speech,conflicting opinions naturally appeared in online comments on the global Internet. Offensive language or vulgar language was used to vent emotions or express extremely supportive or negative views under fierce disputes.

6.2 Cluster Analysis of the Video Comments

Regarding to the design of the experiment, We conducted multiple experiments with the number of clusters from 2 to 15, and found that when the number of clusters was 10, the results were the best (the average distance of the clusters is small and the clusters can be seen very clearly, which means the language characteristics of comments in this cluster can be recognized easily). As shown in the figure 2, we use a two-dimensional scatter plot of each cluster. Different clusters are represented by different colors and their labels are defined as 0-9.

Through the cluster analysis of 27,324 comments from September 2019 to December 2020, we got 10 clusters. We observe manually on random 1000 comments in each cluster. On the basis of the three language characteristics, we found that each group has their own independent characteristics. Let us discuss them one by one.

There are a total of 3432 comments in cluster 0 (the green area above). The sentence structure of the comments is mainly short sentences, and emojis are sometimes added. The content of the comments is generally based on a positive attitude. For example congratulations on Greta's birthday and agreement of Greta's speech, such as "A man who trusted you", "Wow amazing!!!!!!". But on the other hand, there are many repetitive comments expressing positive attitudes in the comments. For example, we found at least 20 "well said". Therefore, we do not rule out the possibility of increasing positive comments deliberately.

There are 3768 comments in cluster 1 (the purple area below), and the sentence structure of the comments is dominated by long
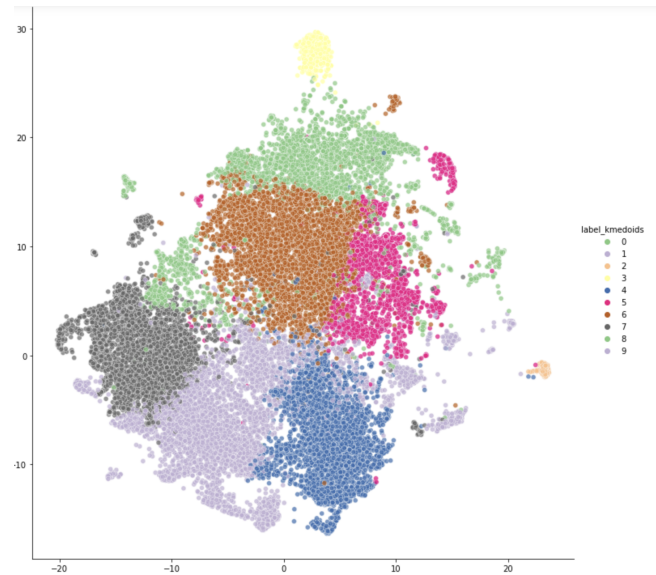

Figure 2: 10 clusters of the video comments

paragraphs. The comments here do not simply express emotions or attitudes, but express one's own opinions and will clarify the reasons or give examples. The content of the comment is not just a response to Greta or her Speech. Comments are more about discussing climate, political ideology, and historical processes from a macro perspective. For example, "Mark my words as History student- she is going to be most influential people in the 21st century like we have seen mahatma gandhi,FDR,Kennedy,mandela,martin luther king in 20th century. we are witnessing history of the future mainstream political leader like greta,malala etc.... so on, pls dont criticize the visionary leader like greta. "We can see that these comments are more rational, and their attitude towards Greta is also vague. The long arguments also greatly increases the number of replies to these comments.

There are 198 comments in cluster 2 (orange area on the right), and the comments are mainly in German. There are many comments on the content that are not relevant to our topic. Therefore, we believe that this cluster is not worth analyzing.

There are 2025 comments in cluster 3 (bright yellow, at the top of the scatter plot), most of which are emojis or individual words that express emotions. In addition, there are a few meaningless characters and robot comments. This clustering has appeared in many experiments. It is well clustered on the top of the scatter plot and there is no intersection with other clusters. We can know that this kind of comments are just some meaningless irrelevant characters and emojis, which are not helpful to our analysis and can be ignored. We call this "invalid comment" cluster.

There are 3644 comments in cluster 4 (the blue area below), and the sentence structure of the comments is mainly long paragraphs. Most of the comments on the content fully agree with Greta's views and are very concerned about the topic of environmental protection. For example "I love you Greta and I'm sorry that you have to live like this. We have to worry about our future on an environmental. It's making me sick, every second is being wasted. The system needs

to change, we need regulations. We also need Efficient Leaders with sympathy. You are amazing, we will keep fighting for our planet!" Most of their comments have strong emotions.

There are 3101 comments in cluster 5 (the rose red area on the left), and the sentence structure of the comments is mainly short sentences and emoticons. The key word for this cluster is "How dare you". Most of the comments were jokes created from these key words, and they were all making fun of Greta. Obviously, this kind of comments does not care about Greta's speech or environmental protection. For example, "Ok guys the how dare you drinking game", "Me: hello Greta: HOW DARE YOU", "new drinking game: everytime Greta says how dare you, take a shot".

There are 4,525 comments in cluster 6 (the central brown area), and the sentence structure of the comments is mainly short sentences. The key word for this clustering is "believe". Most of the comments were distrust of Greta and attacks on the supporters behind her. For example, "you can smell their lies from thousand kilometers away...", "lol. Do you seriously believe that?" They believe Greta is a kid who has been exploited and will also attack Greta. Some comments expressed concern for environmental protection. For example "We all are a part of climate change lol just you love what started must end the earth."

There are a total of 3,914 comments in cluster 7 (the gray area at the bottom left), and the sentence structure of the comments is mainly short sentences or paragraphs. This cluster is similar to cluster 6. The comments are negative about Greta's views and believe that Greta is being used by others. The difference is that the comments in this cluster are full of sympathy for Greta and think Greta is innocent. For example, "Oh dear, she just seems full of obsessive hatred-hope she gets better soon.","Something ruined this girls childhood..."

There are 878 comments in cluster 8 (the green area on the left), most of which are about Greta's behavior. The key word in this cluster is "acting". Most people think Greta's behavior is "bad acting", such as "Very poor acting skills", "Worst acting ever". In addition, there are very few people who think this is "good acting", but after that a laughing emoji is added, such as "A great actor.. hahaha", "You can be a great actor one day!" We can think of it as A kind of irony. We can conclude that almost all comments in this cluster are negative comments on Greta's behavior, which we call the "bad acting" cluster. We can classify it as a "negative attitude" in the following statistics.

There are 1814 comments in cluster 9 (the gray area on the right), and the sentence structure of the comments is mainly short sentences. The key word for this cluster is "children". The commentary believes that environmental protection is not as necessary as saving poverty and hunger, and at the same time expressed concerns about children in war. They believe that Greta's discussion is unnecessary, because there are many more serious problems that we need to solve. For example, "What about the kids in syria, gaza, libya and yemen?". In the end, the commentary moved towards the discussion of poverty in Africa, which completely deviated from our topic.

6.3 Attitude Analysis

Through combing and analyzing the data of 27,324 online comments (25,101 of which are valid comments) generated from a video

Table 4: Attitude distribution

Topic	friendly	negative	vague
Greta	Cluster 0,4,7	Cluster 5,6,8	Cluster 1,9
Environmental Protection	Cluster 1,4,6,7	Cluster 9	Cluster 0,5,8

of Greta's speech on the YouTube website between September 2019 and December 2020, according to their language expression characteristics ,content From and attitudet, this study summarizes six major trends in online attitudes, namely, attitudes towards Greta and environmental protection topic: friendly, negative, and vague. "Friendly" refers to the comments and opinions expressed to support Greta or to express concern about environmental protection. "Negative" refers to negative, accusatory, and unfriendly views of Greta or the environmental protection. "vague" means that the subject and content of the review do not show obvious positive or negative views, or the content of the review is irrelevant to environmental issues, and its attitude cannot be clearly defined. Such attitudes often appear in responses to comments made by a certain netizen, that is, "comments of comments". The attitudes expressed may not be the attitudes in line with the subject, but the attitudes or opinions on a certain comment. In addition, 2223 comments in the comment data are invalid data, which refers to German comments and meaningless emoticons or characters. Because German comments are classified by language, they are mixed with six attitude tendencies, which cannot be counted. Next, this section will perform descriptive statistical analysis for specific trends, and visualize statistical results to support the results of qualitative analysis.

According to Appraisal Theory[18], we classify each cluster(as table 4 shows) and count the number of comments in each category to get a pie chart.

From table 4 we know that the clusters which support both Greta and environmental protection are 4 and 7. There is no cluster that support Greta but against environmental protection(we have already guessed it). Only cluster 6 support environmental protection but against Greta. And there is no cluster supports neither. Here we only count the clusters with obvious attitude tendencies. We classify other clusters with fuzzy attitudes as "unclear".

7 DISCUSSION

In this section, the research questions were studied and answered based on the result from both qualitative and quantitative section. Moreover, the coherence and the difference between both analysis methods were discussed.

Based on the result obtained from qualitative analysis and the quantitative analysis, the research questions could be answered with supporting data. The first question 1 is on the audiences' attitude towards Greta Thunberg. In the qualitative section, we have observed both supportive and opposite comments and with the help of machine learning, this question was further studied qualitatively. 43.8% comments show positive attitude, 33.9% negative while 22.3% remain neutral or has not expressed their attitude(fugure 3).

Analogous to the first research question, second research question focused on the attitude towards environmental protection,
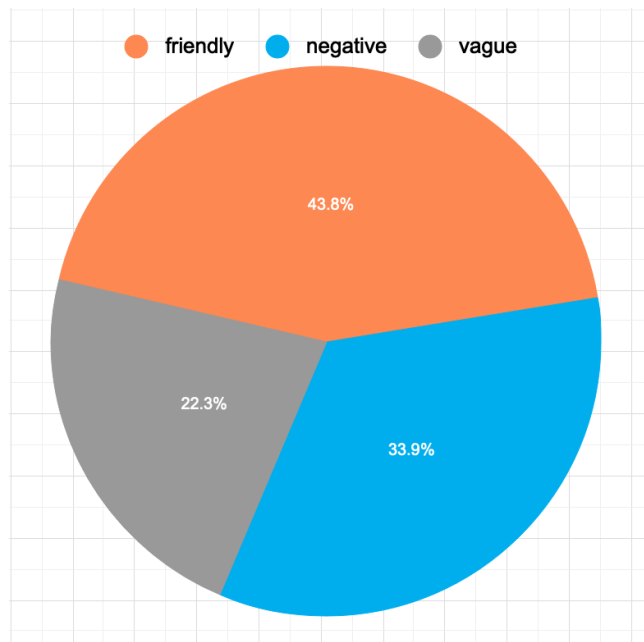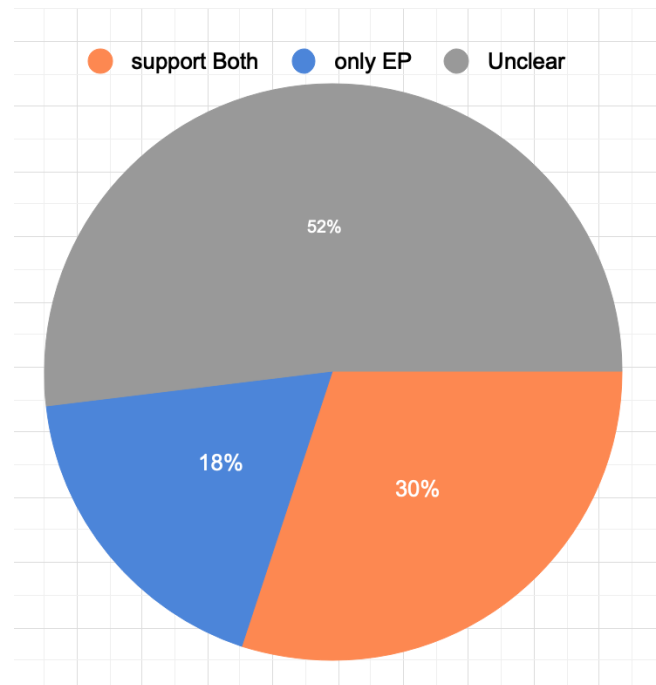
Figure 3: Attitude towards Greta
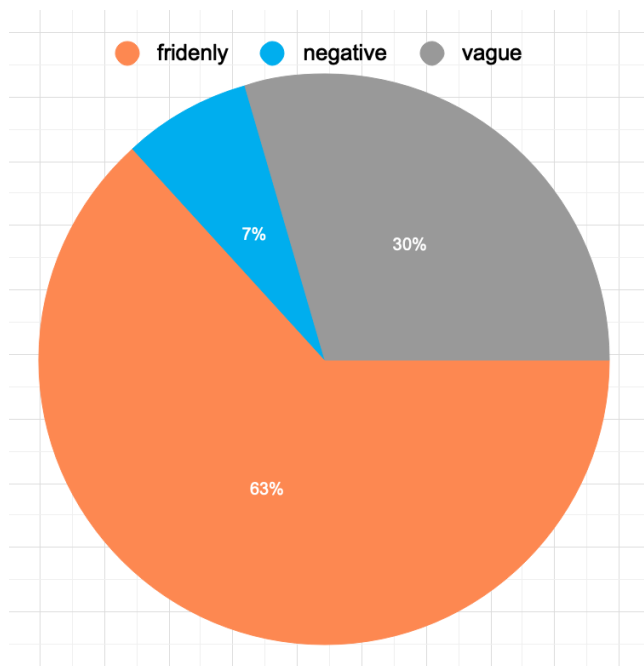

Figure 5: overlap of both attitudes

Figure 4: Attitude towards environmental protection

the comments support environmental protection but oppose Greta. The remaining 52% are vague about at least one aspect of these two.

In summary, the result of qualitative and quantitative analysis accord with each other. The quantitative analysis used machine learning technique to futher refined the classifications mentioned in the qualitative part and broaden the aspects to the comments analysis.

The research goals were also fulfilled with supporting data and scientific analysis. The speech by Greta Thunberg successfully drawn people's attention for climate change and among the comments related to the environmental protection, 90% voiced support.

The second goal was to investigate the attitude of the audiences towards Thunberg's speech. It is delightful to see that most of people were impressed by her speech, especially the hit line "How dare you?". However, according to the result from the qualitative analysis, majority of these comments misunderstood or misinterpreted this sentence on purpose to tease her overacting and the way she gave this speech. So, in one word, the content of her speech has successfully raised audiences' concern on the climate change and expressed her attitude toward EU and world leaders, but at the meanwhile, the way of the speech is unfavorable or unacceptable for most of the audiences.

The third research goals are concerning the speech content. According to the result of speech transcript analysis, it is no wonder that the sentence "How dare you?" is the hit line and aroused heated discussion in the comments. Besides it, the discussion about the childhood and study inspired by "You have stolen my dreams and my childhood with your empty words." and "I should be back in school" also appeared for a certain amount.

which is the one of the core topics in the speech. The results witnesses 63% supportive rate, 7% opposite rate and 30% neutral or not answered(figure 4).

According to Figure 5, we can conclude that 30% of the comments support both Greta and environmental protection. 18% of

In the qualitative analysis, sentiment and attitude analysis was directly applied to the comments with the help of appraisal theory, and in the quantitative section, the attitude expressed in each subset of the comments were also studied according to the appraisal theory. Both analysis methods include the manual work. In comparison to the qualitative analysis, the result obtained from quantitative analysis represents the data better as it divided the positive and negative comments further into six subgroups which show distinct attitude toward Environmental protection and the speaker Greta Thunberg. The classification without help of machine learning has proven to be incomprehensive and incomplete, but it is necessary for the ambiguous dataset (here the comments show vague attitude) to be double-checked and correctly classified.

As Baumer et al. have reported, the machine learning method can deal with massive datasets, which is time-consuming for the existing researcher practices (Baumer, 2017)[4]. On the contrary, there are still several textual information not being able to rely on knowledge about the backgrounds of the topic or being able, e.g. nuances and irony identification. This work indicates the compatibility and complementarity of machine learning techniques and the empirical work. The final result accords with the initial hypothesis.

8 CONCLUSION

This research analyzes a Speech on protecting the environment by Greta Thunberg on YouTube. Firstly, we conducted the critical discourse analysis on the speech content and explored the discourse strategies; secondly, we conducted a qualitative analysis of 27,324 online comments and summarized three main language features, namely, colloquialism and emoticons, The weakening of grammar and the diversity of sentence structure, as well as the frequent use of swear words and offensive language to express extreme emotions; finally, through machine learning clustering models and descriptive statistics, we have drawn six categories of attitudes towards Greta and environmental protection expressed from the comments, Namely "friendly", "negative" and "vague". All in all, this research provides new research perspectives for further understanding of Greta Thunberg and environmental protection, and makes a useful attempt for future social media analysis.

9 LIMIT AND FUTURE WORK

In the original study, the three-dimensional framework of critical discourse analysis by Fairclough (1995)[10] was mainly used in talk shows to discover the discourse strategies, power relations and ideological conflicts hidden between the participants in the dialogue. But our data is a speech, and there is no competition for the right to speak. Therefore, the three-dimensional framework in the original research cannot be directly applied to our research. In response to this problem, we propose two hypotheses.

- A speech is essentially a dialogue between the speaker and the audience. We can assist the analysis of the text by adding changes of the audience's attitude.
- Add a new CDA method, such as systemic functional linguistics.

Besides, the comments of the chosen video contain some special characters, including emoji or some garbled text, which cannot be filtered out efficiently by our current data cleansing process.

For a more scientific and clear research result, we will consider the pattern and characteristics of these undesired content, and add extra criteria to filter them out.

Two further work could be done as a complement this work. First of all, a new dimension, i.e. time scale could be introduced to the research the impact of the major events by Great Thunberg on the audience attitude toward both herself and environment protection. The fundamental work for this research has been completed and the major events were selected based on the amount of audiences of the participants. However, this research question was not investigated due to the lack of time.

Secondly, framing analysis could be applied to the speech transcript and the comments in this work. It is a theoretical approach that has been widely used in communication studies, politics and social movements (Nelson, Oxley, Clawson, 1997)[16]. With the help of frame theory, the topic involved in the comments and the speck transcript could be analyzed. The overlap ratio of the frames in the speech and comments would be of great interest.

REFERENCES

[1] 2014. YouTube API. Retrieved October 7, 2014 from https://developers.google.com/youtube
[2] 2020. Pipeline of CSMA FUB-HCC. Retrieved January 11, 2021 from https://github.com/FUB-HCC/seminar_critical-social-media-analysis/tree/master/Pipeline
[3] 2021. YouTube Data Tool. Retrieved January 11, 2021 from https://tools.digitalmethods.net/netvizz/youtube/
[4] Eric PS Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (2017), 1397–1410.
[5] BBC. 2019. The Swedish Environmental Girl: Greta Thunberg at the center of the controversy. Retrieved October 5, 2019 from https://www.bbc.com/ukchina/simp/49892803
[6] Allan Bell and Peter Donald Garrett. 1998. *Approaches to media discourse.* Wiley-Blackwell.
[7] Danah M Boyd and Nicole B Ellison. 2007. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication* 13, 1 (2007), 210–230.
[8] R Brown and A Gilman. 1960. The Pronouns of Power and Solidarity. En Sebeok, TA (Ed.), Style in Language (pp. 253–276).
[9] Connect4Climate. 2018. Greta Thunberg full speech at UN Climate Change COP24 Conference. Retrieved December 15, 2018 from https://www.youtube.com/watch?v=VFkQSGyeCWg
[10] Norman Fairclough. 2013. *Critical discourse analysis: The critical study of language.* Routledge.
[11] Richard Hanna, Andrew Rohm, and Victoria L Crittenden. 2011. We're all connected: The power of the social media ecosystem. *Business horizons* 54, 3 (2011), 265–273.
[12] Simon David Hirsbrunner. 2021. Negotiating the Data Deluge on YouTube: practices of knowledge appropriation and articulated ambiguity around visual scenarios of sea-level rise futures. *Frontiers in Communication* 6 (2021), 11.
[13] Geoffrey N Leech. 2014. *A linguistic guide to English poetry.* Vol. 4. Routledge.
[14] Prem S Mann. 2007. *Introductory statistics.* John Wiley & Sons.
[15] James R Martin and Peter R White. 2003. *The language of evaluation.* Vol. 2. Springer.
[16] Jack L Nelson and Kenneth Carlson. 1972. Thomas. *Linton.(Eds.) Radical Ideas and the Schools. New York: Holt* (1972).
[17] Muhammad Sipra and Athar Rashid. 2013. Critical discourse analysis of Martin Luther King's speech in socio-political perspective. *Available at SSRN 2928691* (2013).
[18] Peter RR White. 2015. Appraisal theory. *The international encyclopedia of language and social interaction* (2015), 1–7.
[19] Ruth Wodak and Paul Chilton. 2005. *A new agenda in (critical) discourse analysis: theory, methodology and interdisciplinarity.* Vol. 13. John Benjamins Publishing.
[20] YANG Xinzhang. 2010. James R. Martin's Development and Application of Register Theory [J]. *Contemporary Foreign Languages Studies* 10 (2010).
[21] Junchen Zhang. 2018. Critical Discourse Studies (CDS) of Social Media—Taking a Political Interview about China's Issues on YouTube as a Case Study. (2018).

A ORIGINAL TEXT(PROVIDED BY UNITED NATION)

My message is that we'll be watching you?

This is all wrong. I shouldn't be up here. I should be back in school, on the other side of the ocean. Yet you all come to us young people for hope. How dare you!

You have stolen my dreams and my childhood with your empty words. And yet I'm one of the lucky ones. People are suffering. People are dying.

Entire ecosystems are collapsing. We are in the beginning of a mass extinction, and all you can talk about is money, and fairy tales of eternal economic growth. How dare you!

For more than 30 years the science has been crystal clear. How dare you continue to look away, and come here saying that you're doing enough when the politics and solutions needed are still nowhere in sight.

You say you hear us and that you understand the urgency. But no matter how sad and angry I am, I do not want to believe that. Because if you really understood the situation and still kept on failing to act, then you would be evil. And that I refuse to believe. The popular idea of cutting our emissions in half in 10 years only gives us a 50% chance of staying below 1.5 degrees (Celsius) and the risk of setting off irreversible chain reactions beyond human control.

Fifty percent may be acceptable to you. But those numbers do not include tipping points, most feedback loops, additional warming hidden by toxic air pollution or the aspects of equity and climate justice. They also rely on my generation sucking hundreds of billions of tons of your CO_2 out of the air with technologies that barely exist.

So a 50% risk is simply not acceptable to us – we who have to live with the consequences. To have a 67% chance of staying below a 1.5 degrees global temperature rise – the best odds given by the (Intergovernmental Panel on Climate Change) – the world had 420 gigatons of CO_2 left to emit back on Jan. 1, 2018. Today that figure is already down to less than 350 gigatons.

How dare you pretend that this can be solved with just "business as usual" and some technical solutions? With today's emissions levels, that remaining CO_2 budget will be entirely gone within less than eight and a half years.

There will not be any solutions or plans presented in line with these figures here today, because these numbers are too uncomfortable and you are still not mature enough to tell it like it is.

You are failing us. But the young people are starting to understand your betrayal. The eyes of all future generations are upon you. And if you choose to fail us, I say: We will never forgive you.

We will not let you get away with this. Right here, right now is where we draw the line. The world is waking up. And change is coming, whether you like it or not.

Thank you.