Hertta Salokorpi
Computational Literacy
Final Project

Analysis of cult-related language through computational methods

- My academic background is in cultural studies, with a focus on comparative studies of religion. Recently, I have been interested in cults, or what are often referred to as "high-control groups." I am particularly interested in exploring the universal traits of cults, patterns that transcend cultural and historical boundaries. Even without religious context, cult-like groups offer fascinating perspectives for examining psychological aspects and social dynamics of people, and especially the important role of language in shaping how the mind operates. I figured that examining the language of cults could be an interesting way to practice using computational tools.

- A Youtube video called "How Cults Use Language to Control" by the channel *Storied* worked as an inspiration for my project. The video mentions the book "Cultish" (2021), where a linguist named Amanda Montell aims to define common traits of the language used by cults, or more specifically "destructive cults", which are authoritarian groups that harm and exploit their members. Some of these traits are **loaded language** and **jargon.** Loaded language is a term for words and phrases that cause deep emotional reactions, that are utilized by cults in relation to their belief systems. This language is short and dynamic, and purposefully repetitional, which creates efficient and intense emotional associations within cults. Jargon is used to determine who belongs to the group and who does not; it is a kind of shared "secret" terminology, which creates a sense of superiority among group members.

- The aim of this small project is to possibly identify and analyze instances of loaded language or jargon in cult language. However, since I am new to these computational tools

and have no prior experience of them, my primary goal with this project is to familiarize myself with these tools and to test them out, and use this project as a general introduction so that I can possibly take this type of research further in the future.

- The tools that I use are Youtube-Transcript, Python, Excel (find and replace tools) and Voyant Tools (Cirrus).

1. Researching and deciding on the data

Choosing appropriate data proved to be a bit challenging at first; surprisingly, it was difficult to find text suitable for analysis with computational tools. Initially, as I had watched documentaries on different high-control groups, I could empirically notice many similarities between them and their language, even though their belief systems or intentions were not alike. However, finding accessible data that I could compare with each other outside of these documentaries turned out to be more difficult than expected. Ultimately, I decided to use YouTube videos as my source material, transcribing the content into text for analysis, as well as text from a website:

- Father Jones last speech (video) *A speech held by Jim Jones, leader of the People's Temple cult.*

- Heaven's Gate Cult Initiation Tape Part 1 (video) *Initiation tape by Marshall Applewhite, leader of Heaven's gate.*

- Text from the 5D Full Disclosure library *A random segment from a document called "Transitionary Government Ascension guide" in 5D Full Disclosure's website – a cult that used to operate under the name Love Has Won, until their leader Amy Carlson or "Mother God" died, or ascended, as the members believed.*

The data for this project was selected somewhat randomly, but there are similarities between these groups. In summary, while these three cults differ in their belief systems, they share a history of physical and mental abuse, extreme manipulation, and overall toxicity. Both Heaven's Gate and The People's Temple tragically ended in mass suicides, while 5D Full Disclosure/Love Has Won continues to operate today.

As a side note, not all cults are as extreme as the examples I've chosen, but the groups I chose for this project happen to exhibit more extreme outcomes. As Montell, the aforementioned author points out in her book, linguistic tactics related to cults aren't exclusive to cults—they can be seen in various aspects of society, appearing in settings like corporations, politics, and even gym memberships.

2. Pre-processing corpuses

→ I utilized a Free YouTube Transcript Extractor to transcribe the videos to text format. I was a bit suspicious about it at first, but comparing the videos and text they seemed to match and overall, it seemed like a valid, trustworthy tool to me.

→ I removed stopwords using Python using this code:

```
import nltk from nltk.corpus import stopwords

nltk.download('stopwords')

stop_words = set(stopwords.words('english'))

text = "This is an example sentence demonstrating the removal of stopwords."

from nltk.tokenize import word_tokenize nltk.download('punkt') # Required for tokenization words =
word_tokenize(text)

filtered_words = [word for word in words if word.lower() not in stop_words]

print("Original Text:", text) print("Filtered Text:", ' '.join(filtered_words))
```

→ I used Python again to convert all the text to lowercase - the transcribing tool mostly did this already for the videos, but I used this code to do the rest:

```
text = "Example"
lowercase_text = text.lower()
print(lowercase_text)
```

→ I also cleaned punctuation with Python (specifically for the text from the website, because the video transcripts did not have punctuation) with this code:

```
import string
text = "example."
cleaned_text = ''.join(char for char in text if char not in string.punctuation)
print(cleaned_text)
```

→ I also cleaned up some of the People's Temple text in Excel, with the find and replace tool. For some reason, the stopword removal in Python could not fully remove pronouns from this text, which left a lot of "'ll, 're" and other mess in the text, which I removed with Excel. This did not happen with the other texts, though.

3. Visualization and analysis

I used Voyant tools, an online application used for text analysis to visualize the processed texts and find the most common words in the texts, with the tool Cirrus. Each of the three datasets contained 300-400 words after cleaning them. I created word clouds for each text document and began by analyzing the 50 most common words, comparing the word clouds side by side to identify similarities. I then expanded the analysis to include the 200 most common words, as I noticed this provided deeper insights and added more context to the comparisons.

Voyant pointed out the most common words automatically from each word cloud. As I noticed similarities between these lists, I decided to extend the analysis and comparisons by creating my own type categorizations, based on my own observations. No computational tool was used for this part; this further categorization happened naturally, as I examined the word clouds side by side, and instantly noticed similarities. Some categorizations were left outside of this research project, and I only decided to write down the most prominent categories I noticed, but extended type categorization could be done by comparing the word clouds. Furthermore, this type of categorization could be done with different computational tools in the future as well instead of empirical comparison, which I will elaborate on in the final part.

5D Full Disclosure: ego (14); humanity (8); earth (7); moment (6); mind (5)
Peoples Temple: world (6); want (6); life (6); quickly (5); mother (5)
Heaven's Gate**:** human (14); kingdom (10); father (10); planet (7); civilization (7)

Most common words in each word cloud according to Voyant.

- Parental: *mother, father*
- "Universal": *human, humanity, civilization, planet, earth, world, evolutionary*
- Spiritual/religious: *God, Jesus, heaven, energy, creation, doomsday, prayers*
- Related to urgency: u*rgent, quick, quickly, hurry*
- Negative tones: *fear, agony, pain, anguish, impossible, fail, inevitable, problem*
- Related to change: *transform, leave, chance, recreating, refurbished, prepare, mold*

Examples of distinct types of languages I observed, along with random example words from each word cloud.

Though inspecting lists and word clouds already showed clear connections, further classification helped me make clearer observations and confirm similar patterns between the texts and the

language used in them. This made it more concrete, that the distinct types of language I classified all exemplify jargon or loaded language. It was very interesting for me to notice these similar patterns, especially since the data was selected quite randomly and the groups I focused on all have different beliefs and goals, despite sharing some common traits.

For example, the use of parental language, relating to mother or father, was something I did not expect, as I did not consciously choose these groups based on that similarity. However, this ended up being a great example of how cults work; they almost always have a "mother" or "father" figure, even when they are not straightforwardly referred to as such. Cults are authoritarian groups which are based on serving the needs of one person through manipulation.

What was especially interesting to me was the use of negative language, or language related to urgency and change. This was confirmation for my own previous thoughts, that even when cults belief systems are not that related to religion or spiritual ideas, they almost always tend to sell an idea of change, wanting something, or something being wrong with people or the world (since the use of "universal" language, as well). They tend to sell a solution to a problem through symbolic language.

The symbolic language in these three examples is clear as well, which is exemplified in the use of religious or spiritual words. However, I do not think this specific project really works as the best example of this symbolic language; belief systems in cults usually tend to be quite complicated and extend to massive amounts of information, and these datasets were smaller. (Analyzing the symbolic language of a specific cult could be an interesting research question to focus on in the future.)

Overall, even this basic analysis highlights the use of loaded language and jargon used in cults; even randomly selected sets of language already show linguistic patterns designed to grab attention, evoke emotions, and create a sense of belonging. The automatic lists of common words by Voyant/Cirrus was helpful in creating a quick analysis, which I could take further with my own observations.

4. Final thoughts & notes for further research

During the project, I realized how useful computational tools can be for studying the language of cults, because they make it easier to highlight different connections and patterns. Cults often use

symbolic language, but these tools can help spot recurring words or hidden meanings that have deeper significance. These methods can also be useful in analyzing connections between different cults, as well as cults and more mainstream organizations that use similar language. Through these techniques we can better understand the social and psychological aspects of cults and their language, and importantly, this kind of analysis can help identify harmful patterns, and potentially prevent some of the damage cults or manipulative language can cause.

Overall, this project helped me gain understanding of different computational methods. However, reflecting on the process, I've recognized some areas where I would take a different approach or think of more critically, if I were to tackle this research question again:

- **Repetition:** Language used in cults often relies heavily on repetition. This project did not really emphasize that because the datasets were quite small, considering the size of the project. Emphasizing repetition in speech patterns would likely require more data.

- **Stopwords:** Cults tend to use words like *us*, *them*, *because*, *which*, and *that*—a point also highlighted in the referenced YouTube video. This got me thinking afterwards, that stopword removal may not be entirely appropriate in this context, as words typically classified as stopwords can hold significant meaning here. However, I did notice too late in the project that some stopwords like pronouns were not removed during the cleaning process; sometimes they were removed and sometimes not. This probably created some errors in the datasets, and was a lesson in precision for myself. Overall, this is something to consider and be critical of in future research.

- **Numbers in text analysis:** While numbers are often removed as "noise" during preprocessing, I chose to keep them, particularly in the Heaven's Gate data's context, where numbers appeared more frequently. To my understanding, standard pre-processing suggests eliminating numbers, but I realized they could provide valuable insights in this context. From what I have noticed, cults often emphasize numerology as part of their language, using it as a form of loaded language and jargon. Therefore, in this case, numbers are not simply "noise" but rather meaningful data worth analyzing. However, this did not appear that much in this project because of the amount of data.

- *Sentiment analysis* could work as another interesting computational approach for this research topic, as well as *topic modelling*. My research for this project was relatively simple, as word clouds do not offer deeper analysis. However, methods such as topic modelling could offer deeper understanding for these questions, identifying themes and structures.