The Rise of Memes, Trolling, and Overall Negativity on the Internet:

Identifying Linguistic Traits of Contemporary Internet Discourse

Jerrett Longworth

ENC 1102

Professor Cooper

November 26, 2019

The Rise of Memes, Trolling, and Overall Negativity on the Internet: Identifying Linguistic Traits of Contemporary Internet Discourse

The internet within contemporary society has become unlike any other form of social discourse to date. The incredibly fast-paced nature of communication, however, does create the question of whether or not scholarly research is proceeding with the same acceleration. I offer some research and findings here to assist in this ever-expanding conversation.

The knowledge of discourse communities and sponsorship has been an important topic of discussion between rhetorical scholars for many years. Several studies have found that literacy is sponsored by some outside source (Brandt, 2017), and it has been additionally established that individuals acquire language from these sponsors. Within the contemporary social atmosphere of the internet, these same rhetorical principles apply. As of approximately the last 15 years, the internet has become more cynically inclined, reflecting a change in the culture that surrounds the internet. This shift, to some extent, can be attributed to the emergence of modern trolling and meme culture. However, there is not much research on this actual change in internet discourse beginning around the mid- to late-2000's. This new internet discourse is what I will refer to as "contemporary internet discourse" in this research. This present research tries to identify and better understand how this new social phenomena has changed internet discourse into what it is in the present.

Contemporary internet discourse, as I described, is the current state of internet discourse after the mid- to late-2000's, which is important to differentiate from past internet discourse. Major international social media networks such as Twitter (founded in 2006), Facebook (founded in 2004), and YouTube (founded in 2005) have paved the new metaphorical roadways for contemporary internet discourse due to their sensational, graphic-heavy user interfaces. These

social networks have encouraged exponential remix culture, allowing for ideas to spread to large audiences quickly.

For additional context of the research I present, Sabatini and Sarracino (2019) offer recent conjecture of possible effects of using social media. After their own analysis of surveys with 50,000 people, they determined that "the use of [social networking services] is negatively associated to people's propensity to trust strangers, neighbours, and the police." This means that people are less likely to trust others if they use social media more frequently, which is the ongoing current trend of internet discourse. The study – while merely speculation – proposes that more research and analysis must be done of social media and contemporary internet discourse.

Since the internet is an expansive place of discourse, not all of the implications of contemporary internet discourse can be discussed here. Rather, this research focuses on the community of 4chan since it is recognized as a root of modern trolling and meme culture (Nissenbaum, A. & Shifman, L., 2017; Sparby, E., 2017). The primary literacy artifact, this internet forum, is an extremely volatile, however extremely relevant "breeding ground" of the aforementioned trolling and meme culture. 4chan is based around a highly anonymous community where anyone is free to join, however due to its nature, is considered to be ethically and socially forbidden territory for normal internet users. Generally speaking, the content that users post on 4chan are free of many of the ethical limits found in other communities and pose relatively little moderation to its users. The information from research gathered attempts to better understand how literacy sponsorship and public internet-based discourse communities have changed as a result of contemporary trolling and meme culture in the context of current scholarly conversation.

Literature Review

The internet is a prevalent medium for both producing and consuming content in a variety of topics. The freedom that the internet provides as a platform allows anyone of any background to share their passions through discourse with like-minded communities. Once contemporary internet discourse emerged, these communities are arguably "tainted" by meme and trolling culture. However, before understanding the discussion of contemporary internet discourse, a series of reasoning must be established from the universally familiar to the relatively less familiar. First, I will discuss Brant's ideas of literacy sponsorship and John's ideas on discourse communities. From this, a primitive form of internet discourse will be brought to understanding from Mitra and Watts' discussion of internet discourse, which serves as a point of comparison for contemporary internet discourse. To the present day, I will present Nissenbaum, Shifman, and Sparby's research of the 4chan community to finish the progression of the current scholarly discussion.

Literacy Sponsorship and Discourse Communities: Foundations of Commentary

According to Deborah Brant (2017), literacy sponsorship is the idea that organizations can sponsor – or provide access to – literacy. Aside this idea of literacy sponsorship, Brandt discusses that no one becomes literate on their own; rather, all people are dependent on literacy sponsors to provide literacy. The idea of discourse communities lays the foundation for what influence organizations can have on individuals. Within the internet, users are in essence literacy sponsors as well as receivers of that sponsorship simultaneously. Members that participate in internet discourse not only contribute to the ongoing conversation, but they also ingest the content they see online and (intentionally or not) gain some type of literacy as a result. Whether it be a quite literal interpretation of this by learning a new language or by picking up new

colloquialisms from a specialized forum, audience members will take some of the content they see online and reuse it in a different environment. This concept of reusing content (to avoid a tangential understanding of the conversation) will be discussed later. However, with a warning it should be noted that some online sources may offer information in a way that is biased. This bias is what 4chan exemplifies with its relatively unrestricted language and discourse community. That said, literacy sponsorship, while important for discussion, is not the only fundamental lens through which to study this complex environment.

Ann Johns (2017) is another well-known linguist and discusses the idea of discourse communities. She defines the characteristics as essentially the following: a discourse community shares common goals, communicates through the same medium(s), uses the same literary genres, and has a maximum threshold of membership. Each discourse community has their own sets of conventions and rules specific to that community. These conventions, in turn, help define the discourse community with identifiable traits. The discourse community of discussion is the community within 4chan, which research will show its unique traits.

Earlier Internet Discourse: A Baseline for Contemporary Internet Discourse

Though it could be said with some assumption that the internet's social atmosphere is already known to the reader, this would do a disservice to understanding the nature of a discourse community as complex and neoteric such as this one. To understand this nature, I propose to first understand the social climate of the internet in a more primitive state – an internet less influenced by other more recent outstanding social issues such as internet censorship, trolling, and net neutrality. Anandra Mitra and Eric Watts (2002) describe internet discourse in which viral social media networks such as Twitter, Facebook, and Instagram do not

exist. In this more primitive state, core concepts of internet discourse is described in their work. The more contemporary internet discourse, in comparison, will be explored with the archival study and rhetorical analysis of 4chan.

Mitra and Watts (2002) propose that one lens to view internet discourse is with a metaphorical lens of voice. Voice represents the individual, and with it comes the abstraction of voice onto a digital medium. The discourse space that exists within the internet has some unique properties that is unprecedented among any other form of written communication to date. As they put it, the internet has the "potential to flatten hierarchies of power," wherein that the internet has no central location or time, as it contains all conversation from all locations and times simultaneously. These special properties mean the credibility of internet discourse must be judged differently than if the same conversation occurred physically. Rather, internet speakers' credibility is judged based on "emotional and ethical terms" (Mitra & Watts, 2002). Despite the outdated perspective, this serves as a fundamental line of reasoning to understanding contemporary internet discourse.

Rhetorical Velocity and Cultural Capital: The Exponential Nature of the Internet

In completing the necessary context for this research, Ridolfo and DeVoss (2009) propose the concept of "rhetorical velocity," where a content producer considers the use of a particular medium as a part of rhetorical strategy. As they write as examples, one might consider the digital file type or resolution to use for their audience. Essentially, the exponential nature of the internet has caused content producers to consider more than ever how their content will be reused and remixed due to the relative simplicity of remixing other works. With the internet's fast-paced discourse, internet memes play a particularly strong medium for constant remix.

Nissenbaum and Shifman (2017) declare memes as “cultural capital” in modern society. The deeply intertwined nature of internet memes assist in completing their goal of “forming and signifying communal belonging.” Internet memes (and by extension 4chan’s discourse community) capitalize on cultural influence, which is only amplified further when the speed at which today’s internet users can remix content is as staggering as it is.

With the past scholarly conversation understood, attention can now be moved to the center of discussion here: 4chan. 4chan’s social atmosphere is complex due to its entanglement with many facets of contemporary society, so I believe taking incremental steps up to this point was crucial to understanding 4chan’s discourse community.

To succinctly overview the following research, discussion, and conclusion, I propose the following question: What does discourse on 4chan look like today, in 2019, and how does it affect the rest of the internet? In attempting to understand this, I suggested that the underlying foundations of internet discourse be established. Ann Johns (2017) and Deborah Brandt (2017) discuss literacy sponsorship and discourse communities, respectively. These concepts are the overall foundation of understanding internet discourse in general. On the internet, people consume content produced by others, and therefore are to some extent being sponsored by the internet. Additionally, most internet users participate in discourse communities related to their interests. Mitra and Watts (2002) build on this by transitioning from ‘real’ discourse to internet discourse. Since the internet fundamentally changes communication into a continuous conversation and the authenticity of users is no longer bounded by public image, this shifts how credibility is perceived. Once this is established, the focus can now shift to the archival research and rhetorical analysis of 4chan, since 4chan can be considered a source of the meme and trolling that surrounds contemporary internet discourse (Stryker, 2011).

Methods

This research decomposes the use of language on 4chan into discussing both the "what" and "how" of language being used.

For the "what" of language, I conducted archival research using an automated script that counts the frequency of every word used in the selected sample data. The script is outlined in Appendix A and the source code in full is in Appendix B. Since 4chan is such a large community with wildly different sorts of content, I will collect a large amount of it to gather an "average." This average consists of the /b/ (random), /pol/ (politically incorrect), and /adv/ (advice) boards due to their lack of a specific subject. Simply put, one board alone would not be an accurate representation of the community being discussed in this research, but choosing boards other than these could sway the data to specific genres or topics (ie: /v/ (video games), /sp/ (sports), /fit/ (fitness)).

After this data was collected, a subset of this data was chosen at random for the proposed second portion of 4chan's use of language: the "how." This consists of grammatical patterns used within thread posts like punctuation, capitalization, and sentence structure. These patterns will be collected manually by skimming through some portions of the overall data and creating a list of common attributes found in forum posts.

The purpose of this data collection is to show evidence of contemporary internet discourse under the pretense that 4chan is a source of meme and trolling culture (Nissenbaum, A. & Shifman, L., 2017; Sparby, E., 2017). A timeline of the research is as follows: Using the "Save as…" function in Google Chrome, I will collect raw HTML files of various conversation threads. These will be collected from /b/, /pol/, and /adv/ with little regard for the content in them to gather an objective picture of 4chan. A high reply count (over 100 replies) was the only

condition for selecting a thread since they offer examples of more involved discourse. Once collected, the script ran for all of the files gathered and automatically generated the statistics needed for discussion. Examples of the progression from raw data to analysable data can be seen in Appendices C through E. The final aggregated data was observed for trends in words, centered around negative or derogatory language. To complement the quantitative data of 4chan's discourse, qualitative data was also collected. A subset of the original data was observed for the aforementioned grammatical aspects.

While data collection went smoothly, it should be noted that there is some bias in the raw word count. Due to the way the script parsed the raw HTML data, the first and last words of some replies are counted as one continuous word, and therefore do not contribute to a fully accurate counting. Should this script be recreated at another time, this should be accounted for.

Results

The first 20 words of the three generated lists have been provided. There was a total of 93,681 words over 19 selected threads.

data.csv

Word	Occurrences	Percent of Total
the	3137	3.348598
to	2609	2.784983
and	2160	2.305697
a	2126	2.269404
you	2086	2.226706
i	1738	1.855232
of	1406	1.500838
is	1340	1.430386
that	1198	1.278808
it	1023	1.092004
in	1001	1.06852
for	938	1.00127
are	880	0.939358
they	640	0.683169
have	631	0.673562
be	623	0.665023
this	612	0.653281
not	610	0.651146
with	608	0.649011
on	564	0.602043

data-no-common-words.csv

Word	Occurrences	Percent of Total
is	1340	1.430386
are	880	0.939358
it's	386	0.412037
don't	380	0.405632
was	306	0.32664
i'm	275	0.293549
more	249	0.265796
being	230	0.245514
why	218	0.232705
shit	217	0.231637
you're	210	0.224165
fucking	208	0.22203
right	196	0.209221
white	192	0.204951
fuck	174	0.185737
has	157	0.16759
really	155	0.165455
going	154	0.164388
here	153	0.16332
were	147	0.156915

bad-words.csv

Word	Occurrences	Percent of Total
fuck	471	0.50277
shit	311	0.331978
white	262	0.279673
ass	221	0.235907
jew	191	0.203883
boomer	184	0.196411
fag	165	0.17613
retard	108	0.115285
nigg	107	0.114217
trann	68	0.072587
kike	67	0.071519

The following bullet list is a raw list of the qualitative data collected from searching for grammatical patterns in 4chan's discourse.

Trends in Grammar

- No period at the end of sentences
- Some posters use all lowercase, some use proper capitalization and punctuation
- Full capitalization used to emphasize text
- > has two purposes:
    - To quote a previous poster (file5 No.231787380 for example)
    - To emulate a series of events or a conversation (file5 No.231787273, file18 No.21563888 for examples)
- >> Used to reference a previous post
- Most posts are only a sentence or two, but others can be paragraphs worth
- Some posts are just an excerpt from a different post
    - Used to emphasize or make humor (file10 No.813072133)

Discussion

In discussing 4chan and the data, I would like to present two concepts: 4chan's ability to spread memes and bias, and the negative language shown in the collected data.

The Powerhouse of Internet Memes and the Resulting Bias

It is important to understand that 4chan has historically been a source of many extremely popular and viral memes that have since propagated on the internet (Stryker, 2011). For example, lolcats, rickrolling, and Pepe the Frog have all been viral internet memes that were successful in

propagating on the internet due to, if not directly because of, 4chan's community. Pepe the Frog specifically – while having become less relevant on internet culture due to the passing of time itself – has nonetheless become a symbol of liberty and resistance during the 2019 Hong Kong anti-extradition bill protests (Heidi, 2019). The massive cultural influence that 4chan has amassed from its active community has created mainstays in internet culture that have shaped the culture itself into something that could be considered "biased" from 4chan.

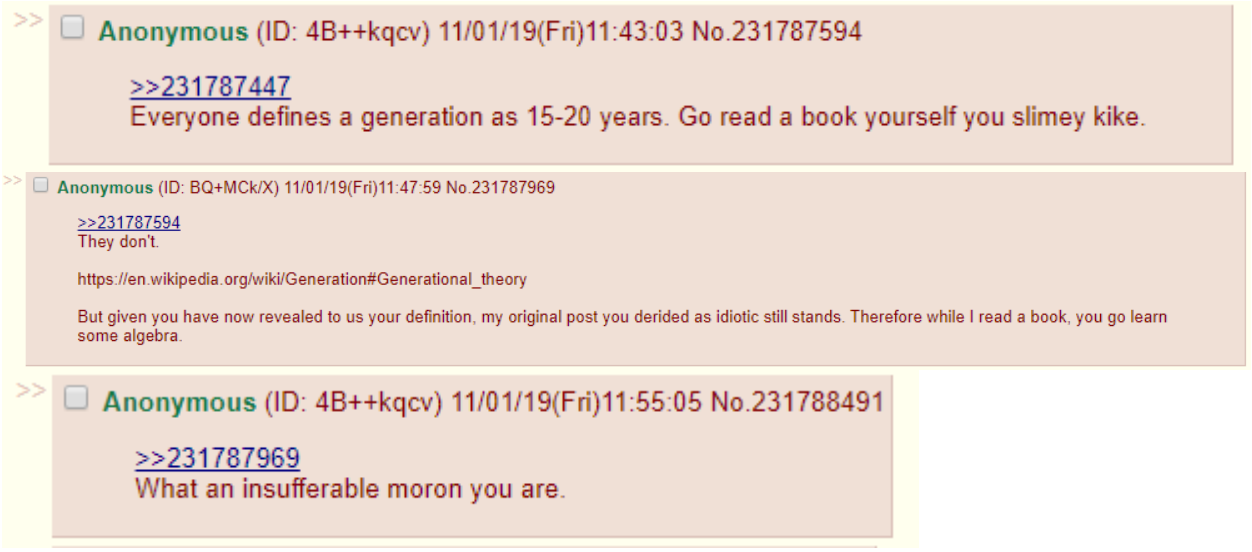
The "bias" that I propose is a direct extension of the ideas that Brant discusses on literacy sponsorship. Recall in this idea that all people are dependent on literacy sponsors to provide literacy. No matter the individual or how self-sufficient they may appear to be, they are unequivocally dependent on others to collect and grow literacy. It should come as a logical conclusion then that the internet demonstrates the same concept through a different medium – a medium that is particularly fast-paced. Essentially those who use the internet are, directly or indirectly, connected to the very creators of any viral internet phenomenon through a series of literacy sponsorships. This literacy sponsorship on the internet is basically amplified *because* the internet itself has become ubiquitous with contemporary culture and plays a role in life for every individual worldwide, no matter if it is direct or indirect.

4chan's community exemplifies an understanding of what Ridolfo and DeVoss (2009) have dubbed "rhetorical velocity," where the use of a particular medium, and the expectation of the *reuse* of the work itself is considered a part of the rhetorical strategy in creating a work. This can be seen from many popular and viral internet memes having an origin at 4chan. Put succinctly, 4chan's community could be considered "in-tune" with internet discourse with an understanding of what specific information can spread on the internet quickly and which method is best for that information to spread.

Returning to the bias on the internet that I propose 4chan presents, the data collected from my research shows that generally, members of 4chan's discourse community use negative or derogatory language simply as a part of how that community functions.

4chan's Negative Language as the Community Norm

The following three replies in an argument of "what millennials are" on /pol/ comprise an extremely small excerpt of the entire conversation. Only an excerpt is shown to maintain a resemblance of a concise discussion.



In this conversation, there are two unique posters (the generated anonymous ID being the only form of identification) that use negative language directed at each other seemingly only for the sake of continuing argument. It should be noted that there are many other replies between these shown here, indicating that within the "chaos" that is seemingly present, there are continuous streams of conversation within the larger discussion. This negative language shown is highly representative of the rest of the overall discussion in this thread, and this thread being representative of /pol/ as a whole. However, this thread is not necessarily just an argument in this context, but is rather an example of a normal discussion taken within the premises of a certain

discourse community. This community discusses in such a way to deter outsiders from entering the community (Nissenbaum & Shifman, 2017).

The following excerpt of a conversation from /b/ contains the start of a thread and two following replies.

A conversation like this, while appearing to the average outsider as incredibly insensitive or misogynistic, is simply part of the normal discourse on 4chan. However, as explained best by the subtitle under 4chan's /b/ heading, "The stories and information here are artistic works of fiction and falsehood. Only a fool would take anything posted here as fact." This gives users the freedom to essentially say whatever they want without repercussions, not only from a public, anonymous standpoint, but from 4chan's standpoint as well. From this, it should be no surprise that conversations like the ones shown can begin to represent the community as a whole.

The negative or derogatory words used within the collected data amounts to about 2.3% of the total word count, and should be considered significant even if it appears to be a small number. Simply the top 100 most frequently used words in the English language alone account for about 50% of all the words used in the collected raw data. The statistical data shows the prevalence of negative or derogatory language to not be an insignificant amount, simply backing

up the rather simple claim that “people on 4chan are saying mean things.” The implications of this, however, may lead to some important distinctions of contemporary internet discourse.

Conclusion

To conclude with my research, I wish to write backwards from the implications to the initial frameworks. In the discussion, I presented the two ideas that 4chan is a powerful source of memes, and that the 4chan community uses negative or derogatory language regularly as a part of normal discourse. Finally, I can conclude this research with its implications. What I propose is these two concepts, of 4chan’s ability to spread viral internet memes, as well as the negative and derogatory discourse that ensues on 4chan, are connected. Logically, this would make sense since they both center around the same community, but the implications of which are much more cynical, leading to the title of this paper. If 4chan is a major source of memes, and is therefore a major source of many tropes on internet culture, but is simultaneously a place where negative language and trolling is used regularly, what are the implications for the internet’s culture itself? Essentially, there may be hidden undertones under many popular memes on the internet that the public is unknowingly propagating, but nonetheless could have been the hidden brainchild of some anonymous 4chan user. Similar to how Pepe the Frog – an internet meme originating from 4chan – was previously mentioned to be a symbol of freedom and liberty in the 2019 Hong Kong anti-extradition bill protests, what other internet memes or general habits of internet culture may be getting remixed as a result of some 4chan user? The undertones of those internet memes may possibly be much darker or otherwise capitalizing on the internet’s fast-paced culture to push a certain political, social, or economic agenda. Additionally, if the internet is becoming increasingly entangled with the culture outside of the internet, at what point do these possible

undertones appear in daily interactions? These points only continue to be validated when paced through credible scholarly research.

From the contemporary internet discourse of today to the internet discourse of the past, earlier forms of internet discourse show that the ways people assess credibility to be more emotional and possibly more vague compared to print texts or physical discourse. Thus, internet users may find themselves believing discourse to be factual when it is in reality no more than entertainment for posters.

Proceeding further back, recall the concepts of discourse communities and literacy sponsorship. If 4chan is a hub of some internet memes in their earliest form, and all internet users are in essence sponsored by the internet in some way, what prevents people to internalize some of the ideals that have originated from 4chan, even if they are not intended to be serious? I believe as the internet continues to grow in prevalence within cultures worldwide, more people (especially those on websites like 4chan) may learn how to propagate their “dark” opinions to an unknowing general audience through seemingly harmless internet memes.

The implications of how certain discourse communities can influence a general audience specifically through the internet must be further researched to understand how society’s discourse as a whole is transitioning due to the introduction of the Information Age. It is certain that the internet already plays a massive role in societal influence, but exactly what this implies for other aspects of culture remains largely unknown. As an example mentioned in the beginning of research, Sabatini and Sarracino (2019) essentially show that people who use social media are less likely to trust strangers. However, they amount their research to speculation at best, since the current research of this internet discourse has not currently been keeping pace with the exponential nature of contemporary internet discourse.

Although this may seem desolate for scholars to keep pace with contemporary internet discourse, I believe that scholarly conversations and research will quicken in pace alongside the internet given time. I hope my research shines a light into a direction not yet fully understood to better understand how literacy sponsorship and public internet-based discourse communities have changed as a result of contemporary trolling and meme culture.

References

/b/ - Random - 4chan. Retrieved November 24, 2019, from https://boards.4chan.org/b/

Brandt, D. (2017). Sponsors of literacy. In E. Wardle, D. Downs, & J. E. Sullivan III (Ed.), Writing about writing: A college reader (3rd ed., pp. 68-81). Boston, New York: Bedford/St. Martin's.

Heidi. (2019, August 20). Pepe青蛙成逆權運動文宣～網民去信告知原作者，原作者：「Pepe for the people！」. *Presslogic*. Retrieved from https://holiday.presslogic.com/article/141813/pepe%E9%9D%92%E8%9B%99%E6%88%90%E9%80%86%E6%AC%8A%E9%81%8B%E5%8B%95%E6%96%87%E5%AE%A3-%E7%B6%B2%E6%B0%91%E5%8E%BB%E4%BF%A1%E5%91%8A%E7%9F%A5%E5%8E%9F%E4%BD%9C%E8%80%85-%E5%8E%9F%E4%BD%9C%E8%80%85-pepe-for-the-people

Johns, A. M. (2017). Discourse communities and communities of practice: Membership, conflict, and diversity. In E. Wardle, D. Downs, & J. E. Sullivan III (Ed.), Writing about writing: A college reader (3rd ed., pp. 319-326). Boston, New York: Bedford/St. Martin's.

Mitra, A., & Watts, E. (2002). Theorizing cyberspace: The idea of voice applied to the internet discourse. New Media & Society, 4(4), 479-498. doi:10.1177/146144402321466778

Nissenbaum, A., & Shifman, L. (2017). Internet memes as contested cultural capital: The case of 4chan's /b/ board. New Media & Society, 19(4), 483-501. doi:10.1177/1461444815609313

Sabatini, F., & Sarracino, F. (2019). Online social networks and trust. Social Indicators Research, 141(3), 229–260. doi:10.1007/s11205-018-1887-2

Seargeant, P., & Tagg, C. (Eds.). (2014). The language of social media: Identity and community on the Internet. Basingstoke, Hampshire: Palgrave Macmillan.

Sparby, E. M. (2017). Digital social media and aggression: Memetic rhetoric in 4chan's collective identity. Computers and Composition, 45, 85-97. doi:10.1016/j.compcom.2017.06.006

Stryker, C. (2011). Epic win for Anonymous: How 4chan's army conquered the web. Abrams Press.

Appendices

Appendix A: Web Scraper Script Outline

The script is created using the programming language Python 3.8 with the BeautifulSoup library and Python's in-built csv libraries. First, the script scrapes the relevant data out of every HTML file within the ./specimens/ folder. Data is cleaned of most symbols, split by each individual word, then counted of word frequency. After this, the script sorts this word count by most common first and is transferred into CSV format. For the uninitiated, this is simply an easy-to-generate text version of a spreadsheet. The script creates three CSV files: data.csv, data-no-common-words.csv, and bad-words.csv. The first, data.csv, is the total collected word count with no special parameters. data-no-common-words.csv is the same as data.csv, however the script automatically omits the 100 most frequent words in the English language according to an analysis of the Oxford English Corpus in 2011. bad-words.csv, while the name of the file may appear misleading, is simply the collected list of words that I would (from a reader's perspective) consider negative or derogatory language in this environment. Words that include the following terms were collected into this file: "fuck," "shit," "white," "boomer," "fag," "jew," "retard," "nigg," "ass," "trann," and "kike." bad-words.csv was manually merged of duplicate entries (arising from variations of the same word) using Microsoft Excel. Additionally, it should be noted that percentages generated from the script are out of the full collected data and does not change based on the parameters that data-no-common-words.csv or bad-words.csv set.

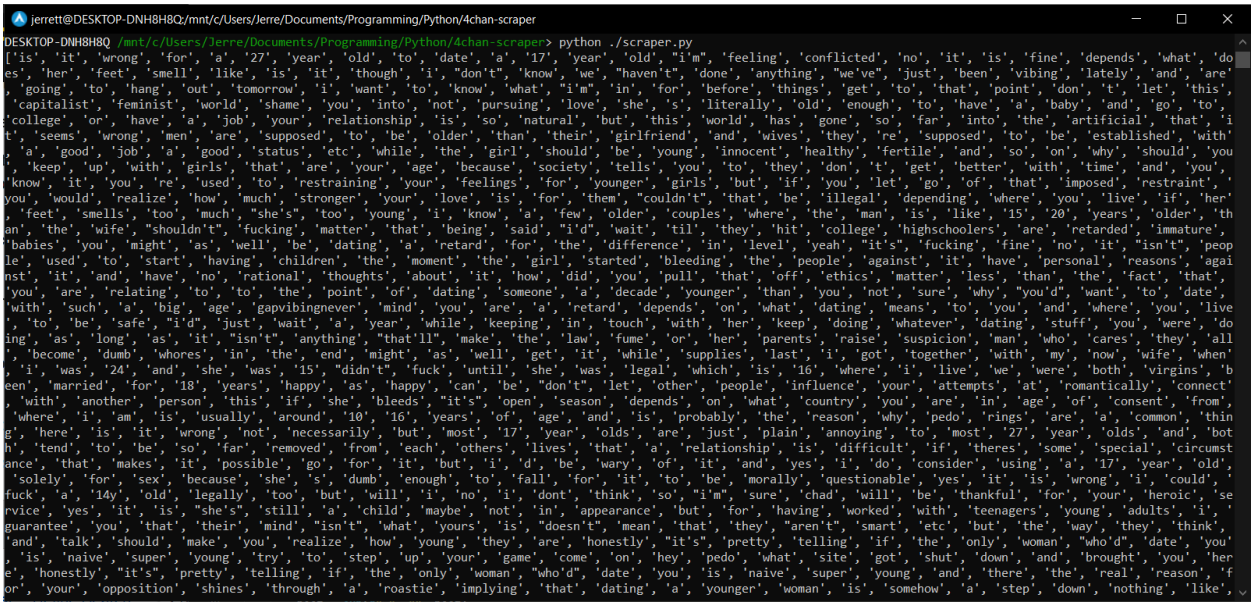
Appendix B: Web Scraper Code

```
#!/usr/bin/python3

from bs4 import BeautifulSoup
from os import listdir
from re import sub, findall
import csv

files = listdir("./specimens/")
files_to_scrape = []
word_list = []
word_freq = {}
total_word_count = 0

most_common_words = ["the", "be", "to", "of", "and", "a", "in", "that", "have", "i", "it",
"for", "not", "on", "with", "he", "as", "you", "do", "at", "this", "but", "his", "by", "from",
"they", "we", "say", "her", "she", "or", "an", "will", "my", "one", "all", "would", "there",
"their", "what", "so", "up", "out", "if", "about", "who", "get", "which", "go", "me", "when",
"make", "can", "like", "time", "no", "just", "him", "know", "take", "people", "into", "year",
"your", "good", "some", "could", "them", "see", "other", "than", "then", "now", "look",
"only", "come", "its", "over", "think", "also", "back", "after", "use", "two", "how", "our",
"work", "first", "well", "way", "even", "new", "want", "because", "any", "these", "give",
"day", "most", "us"]

# Scan for all the html files in the specimens folder.
for file in files:
	if ".html" in file:
		files_to_scrape.append(file)

for file in files_to_scrape:
	try:
		f = open(f"./specimens/{file}", "r")
	except FileNotFoundError:
		continue

	print(f"Scraping file {files_to_scrape.index(file) + 1} / {len(files_to_scrape)}
({file})")

	# Gather a list of all the posts in the html file.
	soup = BeautifulSoup(f.read(), 'html.parser')
	posts = soup.find_all("blockquote", class_="postMessage")

	raw_data = []
	for post in posts:
		# Clean up the posts.
		post = sub(r">>[0-9]{6,10}", "", post.get_text())
		post = sub(r">", "", post)
		post = sub(r"'", "", post)

		# Dump the data into one spot.
		raw_data.append(findall(r"[\w']+", post))
```

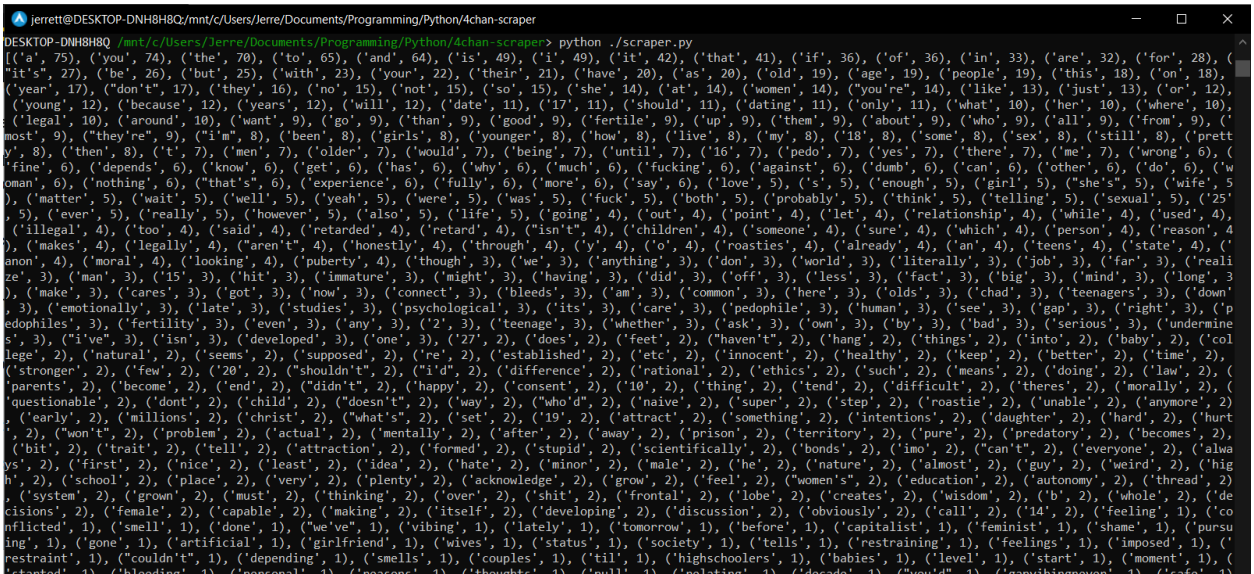
```
	# raw_data is a list of a list of words.
	# This needs to be "flattened" into simply a list of words.
	flattened_data = []
	for sublist in raw_data:
		for item in sublist:
			flattened_data.append(item)

	# Set all the words to lowercase to avoid creating multiple entries for the same word.
	for i in range(len(flattened_data)):
		flattened_data[i] = flattened_data[i].lower()

	# print(flattened_data) Provides Appendix C.

	# Every word is counted then placed into a master word list and a dict with word
frequencies.
	for word in flattened_data:
		total_word_count += 1
		if word not in word_list:
			word_list.append(word)
			word_freq[word] = 1
		else:
			word_freq[word] += 1

print("Data scraped. Exporting to CSV...")

# Sort the word dict and put it in a new array.
sorted_count = [(i, word_freq[i]) for i in sorted(word_freq, key=word_freq.get, reverse=True)]

# print(sorted_count) Provides Appendix D.

# Open a csv file to edit.
try:
	csvfile = open('data.csv', 'w')
except PermissionError:
	print("Permission to edit 'data.csv' denied. Perhaps you have it open in another
program?")
	quit()
csvwriter = csv.writer(csvfile)

try:
	csvfile2 = open('data-no-common-words.csv', 'w')
except PermissionError:
	print("Permission to edit 'data-no-common-words.csv' denied. Perhaps you have it open
in another program?")
	quit()
csvwriter2 = csv.writer(csvfile2)

try:
	csvfile3 = open('bad-words.csv', 'w')
except PermissionError:
	print("Permission to edit 'bad-words.csv' denied. Perhaps you have it open in another
program?")
	quit()
```

```
csvwriter3 = csv.writer(csvfile3)

# Write all the data to a csv file.
for i in sorted_count:
	csvwriter.writerow([i[0], i[1], (i[1] / total_word_count) * 100])
	if i[0] not in most_common_words:
		csvwriter2.writerow([i[0], i[1], (i[1] / total_word_count) * 100])
	if "fuck" in i[0]:
		csvwriter3.writerow(["fuck", i[1], (i[1] / total_word_count) * 100])
	if "shit" in i[0]:
		csvwriter3.writerow(["shit", i[1], (i[1] / total_word_count) * 100])
	if "white" in i[0]:
		csvwriter3.writerow(["white", i[1], (i[1] / total_word_count) * 100])
	if "boomer" in i[0]:
		csvwriter3.writerow(["boomer", i[1], (i[1] / total_word_count) * 100])
	if "fag" in i[0]:
		csvwriter3.writerow(["fag", i[1], (i[1] / total_word_count) * 100])
	if "jew" in i[0]:
		csvwriter3.writerow(["jew", i[1], (i[1] / total_word_count) * 100])
	if "retard" in i[0]:
		csvwriter3.writerow(["retard", i[1], (i[1] / total_word_count) * 100])
	if "nigg" in i[0]:
		csvwriter3.writerow(["nigg", i[1], (i[1] / total_word_count) * 100])
	if "ass" in i[0]:
		csvwriter3.writerow(["ass", i[1], (i[1] / total_word_count) * 100])
	if "trann" in i[0]:
		csvwriter3.writerow(["trann", i[1], (i[1] / total_word_count) * 100])
	if "kike" in i[0]:
		csvwriter3.writerow(["kike", i[1], (i[1] / total_word_count) * 100])

print("Completed successfully.")
```

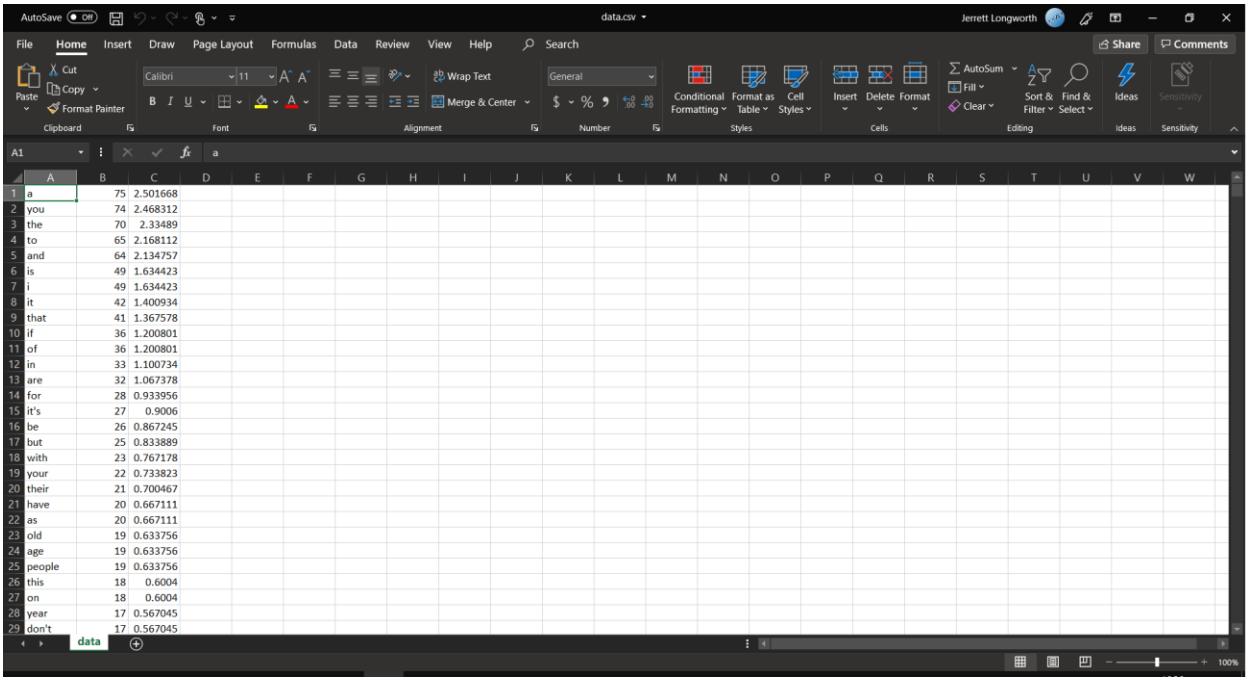
Appendix C: Example of Raw Data (Not Representative of Full Data)



Appendix D: Example of Sorted Data (Not Representative of Full Data)



Appendix E: Example of data.csv (Not Representative of Full Data)



	A	B	C
1	a	75	2.501668
2	you	74	2.468312
3	the	70	2.33489
4	to	65	2.168112
5	and	64	2.134757
6	is	49	1.634423
7	i	49	1.634423
8	it	42	1.400934
9	that	41	1.367578
10	if	36	1.200801
11	of	36	1.200801
12	in	33	1.100734
13	are	32	1.067378
14	for	28	0.933956
15	it's	27	0.9006
16	be	26	0.867245
17	but	25	0.833889
18	with	23	0.767178
19	your	22	0.733823
20	their	21	0.700467
21	have	20	0.667111
22	as	20	0.667111
23	old	19	0.633756
24	age	19	0.633756
25	people	19	0.633756
26	this	18	0.6004
27	on	18	0.6004
28	year	17	0.567045
29	don't	17	0.567045