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

ENC 1102

Esteemed Professor Loren Cooper

November 4, 2019

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

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). 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 "modern internet discourse" in this research. This present research tries to identify and better understand how this new social phenomena has changed internet discourse to what it is in the present.

For supplemental context, 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. The study – while merely speculation – proposes that more research and analysis must be done of social media and modern internet discourse.

Since the internet is an expansive place of discourse, not all of the implications of modern internet discourse can be discussed here. Rather, this research focuses on the community on

4chan since it is recognized as a root of modern trolling and meme culture. The primary literary 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 (which will be discussed later), 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 ethical limits and pose relatively little moderation. The information from research gathered attempts to understand how literacy sponsorship and public internet-based discourse communities have changed as a result of modern trolling and meme culture.

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 modern internet discourse emerged, these communities are arguably "tainted" by meme and trolling culture. However, before understanding the discussion of contemporary internet discourse, a process of reasoning must be established from the universally familiar to the relatively less familiar.

According to Deborah Brandt (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 and corporations (including the internet itself) 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. 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. Literacy sponsorship, although important for discussion, is not the only 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.

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 issues such as internet censorship,

trolling, and net neutrality. The latter portion will be explored with the subsequent archival study and rhetorical analysis of 4chan.

Anandra Mitra and Eric 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 everywhere and all time. 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 modern internet discourse.

In completing the necessary context for this research, 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 as they outline in their discussion (Nissenbaum & Shifman, 2017).

With this in mind, 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 modern society as well, so I believe taking incremental steps up to this point was crucial to understanding 4chan's discourse community.

To succinctly overview the following research, I propose the following question: What does internet discourse look like today, in 2019? 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 in 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 much is established, the focus can now shift to the archival research and then rhetorical analysis of 4chan, since 4chan can be considered a source of the meme and trolling culture that surrounds modern internet discourse.

Research Design

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 will conduct archival research using an automated script that will count the frequency of every word used in the selected sample data. The script is outlined in Appendix A and the source code 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 will consist of the /b/ (random), /pol/ (politics), and /adv/ (advice) boards due to less moderation and 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 other boards could sway the data to specific genres or topics.

Once this data is collected, a subset of this data will be 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 portion of the overall collected data and creating a list of common attributes found in forum posts.

The purpose of this data collection is to show evidence of modern internet discourse discussed in the rest of the paper under the pretense that 4chan is a source of meme and trolling culture. A possible timeline for the research is as follows:

- 1-2 Days: I will use the save function in my web browser to download 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. The script will run for all of the files collected and automatically generate the statistics needed for the final discussion. Examples of the progression from raw data to analysable data can be seen in Appendices C through E.
- 2-3 Days: The final aggregated data will be observed for trends in words, centered around negative or derogatory language.
- 2-3 Days: A subset of the original data will be observed for the aforementioned grammatical aspects. These last two steps may occur simultaneously or sequentially.
 Should the script fail, either in its creation or final use, I will simply use a less interesting method, word count online. I will essentially go to the first web result with word frequency

counting and copy-paste large sections of the page into it. This method is not ideal though, since extraneous information will be collected in the process, like post timestamps, usernames (if applicable), and other website boilerplate.

References

- 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.
- 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.
- Lu, D. (2018). Reddit's diverse platform: Transforming the social media landscape through its communities. *Stylus*, *9*(2), 30-38. Retrieved from https://writingandrhetoric.cah.ucf.edu/stylus/files/9 2/Stylus 9 2 Lu.pdf
- March, E., Marrington, J. (2019). A qualitative analysis of internet trolling. *CyberPsychology*, *Behavior & Social Networking*, 22(3), 192-197. doi:10.1089/cyber.2018.0210
- 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

- Squires, L. (2010). Enregistering internet language. *Language in Society*, *39*(4), 457-492. Retrieved from https://doi.org/10.1017/S0047404510000412
- Wu, T.-Y., & Atkin, D. J. (2018). To comment or not to comment: Examining the influences of anonymity and social support on one's willingness to express in online news discussions.

 New Media & Society, 20(12), 4512-4532. doi:10.1177/1461444818776629

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 to first scrape the relevant data out of the forum, then organize the frequency of this data into CSV format. For the uninitiated, this is simply an easy-to-generate text version of a spreadsheet. This can be imported into Microsoft Excel to generate visual representations of the data to show final results.

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 = []
# files_to_scrape = ["file12.html"]
word_list = []
word_freq = {}
total_word_count = 0
# Scan for all the html files in the specimens folder.
for file in files:
      if (files_to_scrape != []):
             break
      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)
      # 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)
# Open a csv file to edit.
csvfile = open('data.csv', 'w')
csvwriter = csv.writer(csvfile)
# 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])
print("Completed successfully.")
```

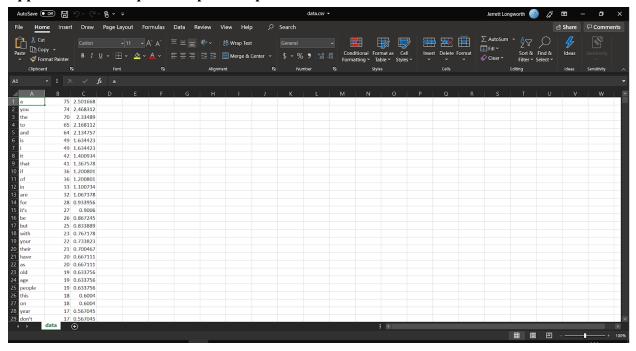
Appendix C: Example of Raw Data



Appendix D: Example of Sorted Data

DEXIDO-DNH8H8Q/mnt/c/Users/lerer/Documents/Programming/Python/Achan-scraper > python ./scraper.py
[f'a, 75), ('you', 74), ('the', 70), ('the', 65), ('ond', 64), ('is', 49), ('it', 49), ('it', 42), ('that', 41), ('if', 36), ('of', 36), ('in', 33), ('are', 32), ('for', 28), ('it', 27), ('be', 63), ('but', 25), ('with', 23), ('your', 22), ('their', 21), ('have', 28), ('old', 19), ('age', 19), ('pople', 19), ('this', 18), ('on', 18), ('your', 12), ('don't', 17), ('they', 16), ('ol', 15), ('on', 15), ('so', 15), ('so', 15), ('so', 15), ('so', 18), ('ol', 19), ('age', 19), ('pople', 19), ('this', 18), ('on', 18), ('your', 12), ('be', 26), ('but', 25), ('will', 12), ('date', 11), ('17', 11), ('should', 11), ('dating', 11), ('only', 11), ('whould', 19), ('about', 19), ('about', 19), ('about', 19), ('be', 19), ('the'', 19), (

Appendix E: Example of Exported Spreadsheet



Appendix F: CITI Training Certification

