

Relationship between the use of swear words and stand-up comedy on YouTube

Conner Hsu

University of California, San Diego

La Jolla, United States

cjhsu@ucsd.edu

Abstract—This paper will investigate the relationship between the usage of swear words and how successful a stand-up comedy video on YouTube is. A sample of roughly 50 videos from the Comedy Central Stand-Up channel was taken. The swearing frequency and like-to-view ratio of each video was calculated. Then, these two quantities were plotted against each other to see if there was a correlation between the like-to-view ratio and the swearing frequency.

The results show that there is a weak positive correlation between swearing and the like-to-view ratio of a stand-up comedy video.

I. INTRODUCTION

Swear words are meant to offend. In Benjamin Bergen's *What the F*, he describes how taboo words can be broken up into four different categories: misused religious words, bodily functions/excrement, sex and sex related acts, and finally slurs; Bergen concisely describes this categorization using the poster childs for each category; hence the "Holy, Fucking, Shit, Nigger" principle [1]. While only slurs are designed to be derogatory words towards certain groups of people, you can easily draw words from each of these categories to form an insult. For words in the "Holy" category, you can "damn" someone or you could call them a "goddamned idiot". For words in the "Fucking" and "Shit" category, there are a plethora of name calling insults that can be drawn from here as well such as "cocksucker" or "shithead".

Despite this reputation for swear words as a tool to offend others, swear words can also make people laugh. As we will see, comedians tend to swear a decent amount. Swear words can be used to make others chuckle just as easily as they can be used to offend. So, it is natural for one to ask if it is worth incorporating swear words into their everyday speech. One might be on the fence about using swear words in their everyday speech. Will incorporating more swearing in their everyday speech make them appear more humorous to their friends and peers? Or will swearing more often just make them come off as brash and rude?

In this paper, the use of swear words in stand-up comedy on YouTube will be investigated in the hopes of generalizing the results to answer the following question: Is there a relationship between the use of swear words in speech and how humorous that speech is perceived to be?

II. PROCEDURE

Data was collected from YouTube's Comedy Central Stand-Up channel. Multiple videos from this channel were analyzed for number of taboo words used and these words were sorted by into the "Holy, Fucking Shit Nigger" categories. A total of roughly 50 videos were drawn videos that were most recently uploaded. After, the transcripts of each video were downloaded as text files and a Python script was used to count the number of taboo words used and categorized each word accordingly. Additionally, the following information will be recorded:

- number of likes and views on the video,
- duration of the video,
- sex of the comedian,
- and number of words from each taboo word category (Holy, Fucking, Shit, Nigger)

With this data, the ratio of likes to views and the taboo word use frequency were computed. Additionally, since the types of taboo words were also recorded, the frequency of each type of taboo word was also computed. The ratio of likes to views are used as a way to measure how successful a particular stand up video was. Finally, the ratio of likes to views can be plotted against the swear word frequency to see if there is a relationship between frequency of swearing and how successful a stand up video is.

The sex of each comedian was also recorded, but not used in the data analysis. A side question that I was curious about was if male comedians tend to swear more than female comedians and if the relationship between swearing frequency and the like-to-view ratio looks different for both sexes.

A. Limitations

Counting swear words is a challenge because the definition of a swear word varies from person to person. Some words that kind of teeter between swear word territory and normal word territory are words like, moron, idiot, crap, and porn. There are also many variants of some swear words and so many such that it is pretty difficult to account for all of them (just think of how many variants that use "fuck" or "shit").

Some words are context dependent. One would argue that saying "Jesus Christ!" is taking the lord's name in vain (or something like that, I am not religious), however it is also perfectly acceptable to just say "Jesus Christ" if it's not being used to exclaim anger or frustration. Accounting for context dependent words requires actually watching all of the videos.

TABLE I
THE AVERAGE SWEARING FREQUENCY FOR EACH CATEGORY OF SWEAR WORDS.

Category	Frequency (words/min)
Religious	0.267
Sex	0.498
Excrement	0.269
Slur	0.134
All	1.297

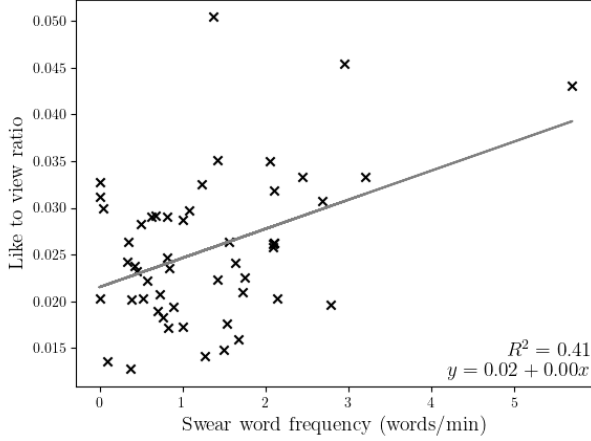


Fig. 1. Like/View ratio with respect to taboo word frequency.

There were also a few videos that actually censored swear words. If a more thorough version of this study were to be done again, it might be worth going back to look at the context of the censored words to figure out what word was actually said.

There is also the question of whether or not the like-to-view ratio is actually a good way to measure success. Do viral videos that have a lot more views have proportionally less likes because the most viewers don't feel the need to press the like button on a video that is already popular? Are viewers less likely to like a video that is not as popular? If any of these are true, some function would need to be applied to the number of likes to account for these effects.

III. RESULTS

The like-to-view ratio was plotted with respect to the frequency of swear words in general and plotted with respect to the frequency of swear words in each category. A line of best fit is also plotted and R^2 value is shown to indicate the strength of the correlation.

IV. DISCUSSION

First thing to notice is that comedians do swear quite a bit. Looking at Fig. 1, only 3 of the videos in the sample featured absolutely no swearing. In fact, Table I shows that on average they swear roughly 1.3 times per minute. This is an indication that swearing is fairly important tool for comedians.

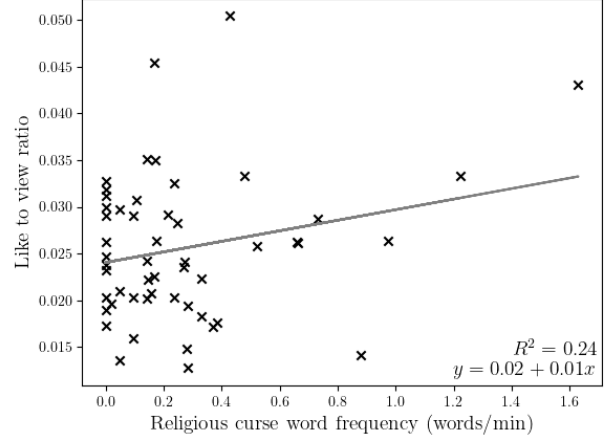


Fig. 2. Like/View ratio with respect to religious curse word frequency.

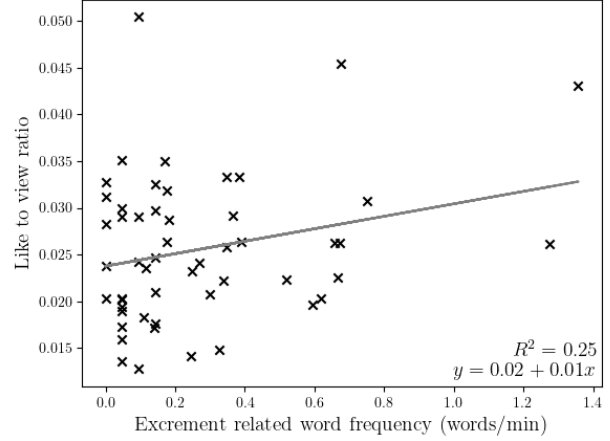


Fig. 3. Like/View ratio with respect to excrement related word frequency.

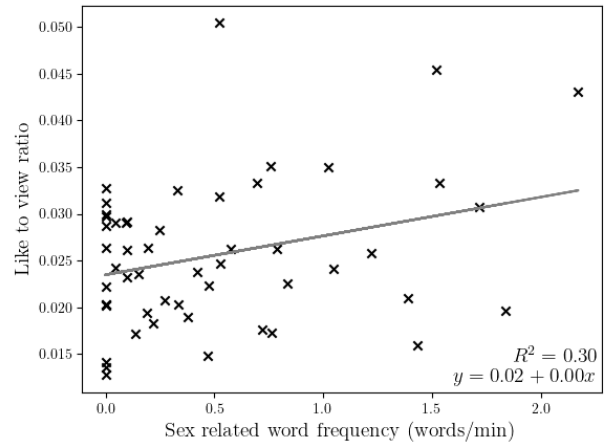


Fig. 4. Like/View ratio with respect to sex related word frequency.

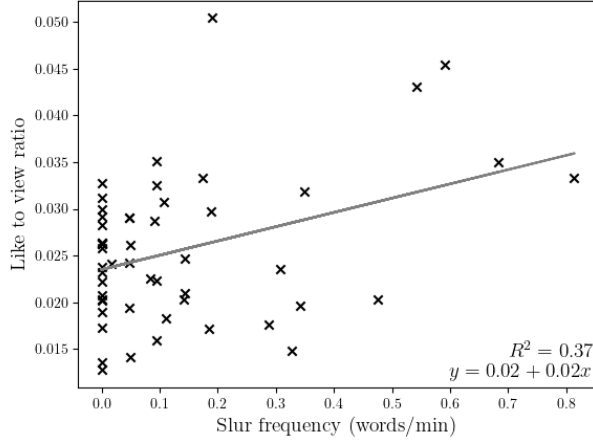


Fig. 5. Like/View ratio with respect to slur frequency.

For all of the figures shown, there is a weak positive correlation between the frequency of swearing and the like-to-view ratio. Despite being used the least on average, slurs have the highest correlation coefficient between using them and the like-to-view ratio. This surprises me because I would think that slurs, out of all the other categories, would tend to offend rather than be humorous. Perhaps, the word's ability to offend makes it more humorous when used correctly.

On the note of using words correctly, just because comedians can swear more to appear more humorous doesn't necessarily mean that *anyone* can swear more to appear more humorous. Of course, one could say that comedians are "good" at swearing as in they know how to use swear words to their advantage instead of coming off as disrespectful and brash.

V. CONCLUSION AND FURTHER RESEARCH

The main takeaway from this study is that there is a weak positive correlation between the use of swear words and the like-to-view ratio of a stand-up comedy video on YouTube. If we assume that the like-to-view ratio is a good metric for measuring how successful a comedian is, then we can also conclude that swearing and how successful a comedian is are positively related.

This then begs the question if swearing more actually causes a comedian to be more successful. This is obviously not necessarily true. Consider an edge case where a comedian's whole bit is just them blurting out "Fuck!" for five minutes. Their swear word frequency would be very high, but I doubt many will find this bit very funny. There are clearly more factors at play. A deeper study would maybe aim to investigate how swear words are used by these comedians to see why a higher swearing frequency would increase success.

APPENDIX

- List of swear words used with categorization
- Source code for data analysis

REFERENCES

- [1] Bergen, B. (2016). What the F: What Swearing Reveals About Our Language, Our Brains, and Ourselves. New York: Basic Books.