Kendall Kaske and Alfred Koo

Final Project

HIST 3368

December 7, 2020

How Language Regarding the LGBTQ+ Community Has Changed

Introduction

Societal views regarding the LGBTQ+ community have been historically negative, but within the last decade, attitudes regarding the community have begun to shift. Over time, our research portrays an increase in accepting language and a decrease in words with a negative connotation. We used data from the Reddit Archive ranging from January of 2009 to May of 2015 in order to research the shift in language. Researching how the LGBTQ+ community is spoken about and how that language has changed over time is critical in understanding the difficulty for the community to be the granted public acceptance that it needs to tackle the large-scale discrimination it is plagued with.

Dataset

The data we used to conduct our research is from the Reddit Archive from January of 2009 to May of 2015. The data was stored on ManeFrame II, a shared high-performance computing environment implemented by Southern Methodist University. The reason we chose the data during this period of time is because the M2 database only has Reddit Archive from October of 2007 to May of 2015. We decided to analyze the most recent six years of data to cut down the size of the data and because we felt those years would be most relevant to the research. As seen in Figure 1, the three main algorithms we are applying to the data are a Tf-Idf analysis for finding the most distinctive words in each year, a Natural Language Processing (NLP) using the spaCy library for finding grammatical gendered and LGBT+ adjective-noun pairs, and found the change of word count for keywords over the years. We went on to provide visualizations and contextualization for all of these findings.

Before applying any analysis method to our data, we targeted Reddit posts regarding LGBTQ+ topics by filtering for posts under sub-reddit "r/lgbt." This process is done by

extracting all the posts under sub-reddit "r/lgbt" and concatenating them into one large string for

every month in a given year. We then grouped the twelve large strings into a panda data frame

and did this for every year from 2009 to 2015. This gave us six different data frames, each with

twelve rows of strings. Each row represents a month of sub-reddit "r/lgbt" posts. During the

```
         Reddit Archive (M2)
           (2009~2015)
                 |
                 v
            Sub-Reddit
             (r/lgbt)
                 |           Lowercase, punctuation & stopwords
                 |               removal, lemmetization
                 v
          Processed Text
        (as Dataframes by year)
    Tf-idf       |  Spacy
    analysis     |   NLP
      /          v          \
  Tf-Idf tables   Parts-of-Speech    Keywords
                                     wordcount
      \          |          /
       \         v         /
          Visualizations
```
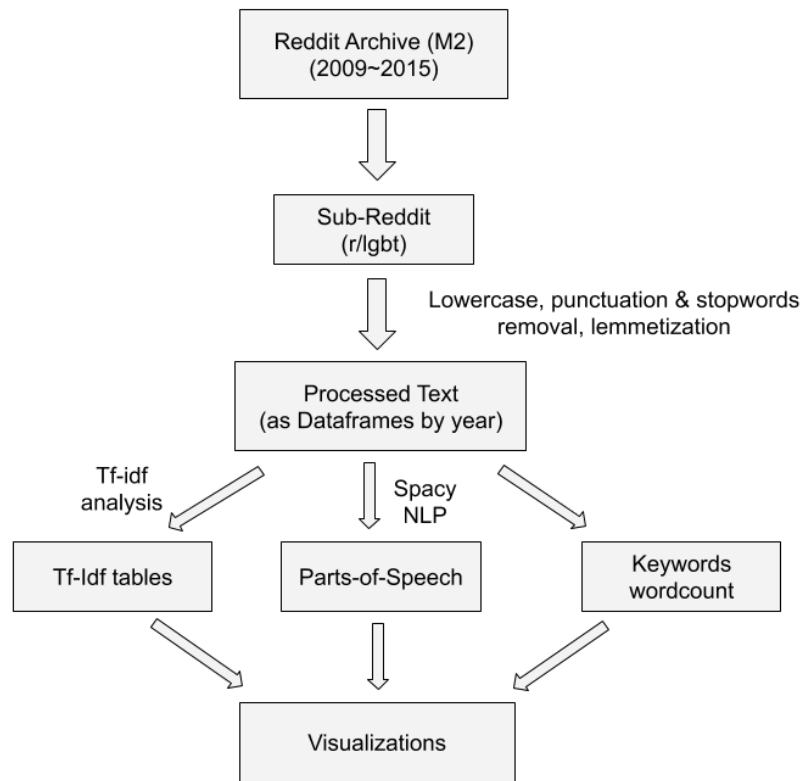
Figure 1: Workflow

filtering process, we also cleaned the data up by doing lowercasing, punctuation removal, stop

words removal, and lemmatization.

For the first analysis method, we did the Tf-Idf analysis on the 2015 and 2010 data to find

the most distinctive words in those years and to evaluate how special words changed in five

years. We made the Tf-Idf tables for those two years, contextualized the top distinctive words,

and visualized the change of word counts for the top distinctive words in the year they appeared

in. For the second analysis method, we did the Natural Language Processing (NLP) using the

spaCy library to find the top grammatical gendered adjective-noun pairs and LGBTQ+ adjective-noun pairs. We identified the top adjective- noun pairs so that we could analyze any trends in language being used. This process is done by using spaCy to extract grammatical parts-of-speech on the 2015 and 2009 data and filter for adjective-noun pairs. For the "noun" in the adjective-noun pairs, we used the words "woman," "man," "lesbian," and "gay." We then visualized all the top grammatical adjective-noun pairs with these four different nouns for both 2015 and 2009. For the third analysis method, we came up with a list of keywords regarding the LGBTQ+ community such as "gay," "lesbian," "trans," bisexual," "man," and "woman." We then found the change of word count of these keywords over the years from 2010 to 2015. In order to properly analyze the importance of the wordcount, we then found the change of word count of these keywords over the twelve months in each given year. We then visualized them to observe the LGBTQ+ keyword trends.

Findings

Between the years of 2009 and 2015, many changes occurred for the LGBTQ+ community, including several comprehensive legislative wins in the United States that were extremely beneficial to the community. These legal changes appear to have had an influence on the dialogue used when discussing the LGBTQ+ community. Legislative advancements were not the only catalyst for the change in dialogue, as current events regarding specific individuals also influenced spikes in wordcount. The overall shift in language that we identified from 2009 to 2015 relied on both legislative advancements and public acceptance of the community. Although

public acceptance of the LGBTQ+ community had significantly increased in 2009 compared to the 1990s and early 2000s, the language being used continued to be overwhelmingly negative. [1]

As a general base, we started our research by compiling key words like gay, lesbian, trans, bisexual, man, and woman to analyze the change of wordcount that occurred between 2010 and 2015. As portrayed in Figure 2, all six of these categories experienced a significant increase
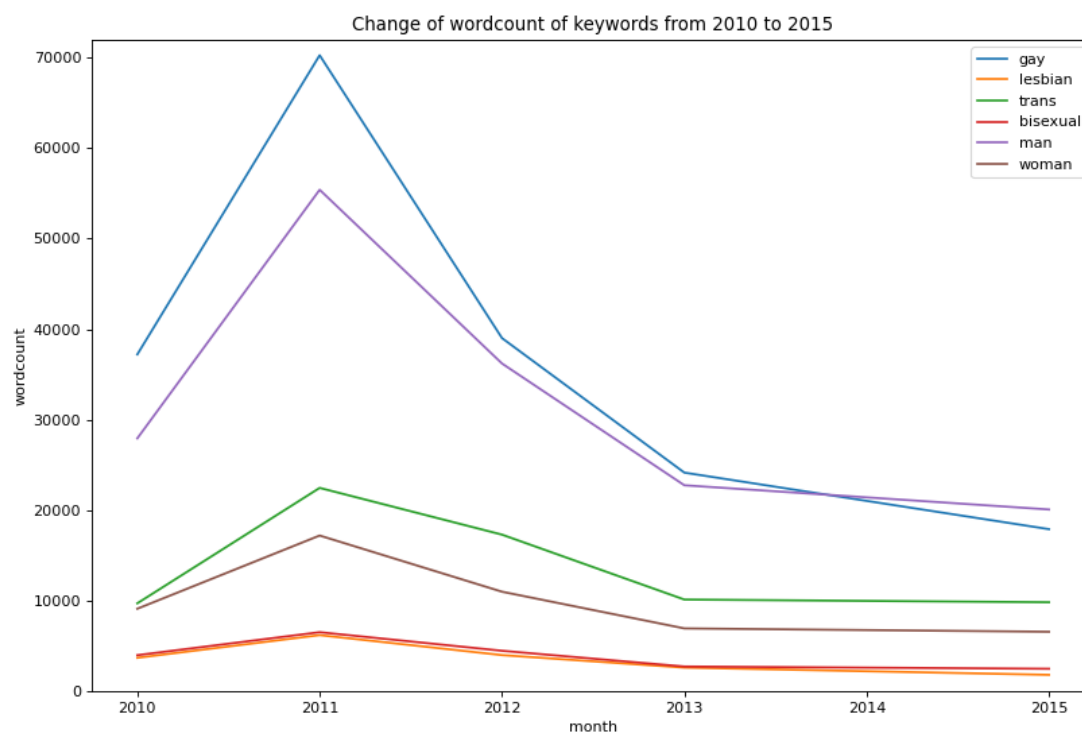

Figure 2: Wordcount of Gay, Lesbian, Trans, Bisexual, Man and Woman

of usage in 2011, which could have been influenced by a major change in legislation within the United States' government. In September of 2011, Congress passed legislation that repealed the Don't Ask Don't Tell policy that had forced LGBTQ+ members of the military to hide their sexual orientation.[2] The ending of this Clinton era policy not only allowed for individuals in the

---

[1] Smith, Tom W., Jaesok Son, and Jibum Kim. "Public attitudes toward homosexuality and gay rights across time and countries." 2014.

[2] Burelli, David. ""Don't Ask, Don't Tell": The Law and Military Policy on Same-Sex Behavior." *Congressional Research Service*, 2010.

military to serve openly, but also helped create a monumental spike in dialogue regarding the LGBTQ+ community. Specifically, there was a major increase of the words "gay" and "man" compared to the other key words that we used. Considering women only made up "14.5 percent of the DoD Active Duty force, while men comprise 85.5 percent of the DoD Active Duty force"[3] in 2011, it is not surprising that the new legislation would create a bigger spike for the words "gay" and "man." As shown in the graph, the wordcount for "gay" went from about 38,000 in 2010 to almost 70,000 in 2011. Similarly, the wordcount for "man" in reference with LGBTQ+ rose from about 28,000 in 2010 to over 55,000 in 2011. The combination of the military policy reform and the spike in wordcount in 2011 portrays an important connection between current events having a substantial impact on dialogue. Although our first word count analysis doesn't specifically express how the language regarding the LGBTQ+ community has changed, we found that analyzing those key words was an important first step.

After analyzing the change in wordcount of general terms, we went on to research the most frequently used words that had a specific relation to the LGBTQ+ community. We found that many of these top words had significant meaning behind them, especially having to do with tragic events or discrimination. There was also notable change in connotation behind the top words in 2009 and 2015. In previous years, many of the top words surrounded incidents that were extremely negative and did not facilitate substantial change. Although some of the top words in 2015 still had a negative connection, a handful had productive endings. In other words, by 2015 language regarding the LGBTQ+ community had begun to shift towards a more positive connotation.

---

[3] Department of Defense. "2011 Demographics, Profile of the Military Community," 2011

As seen on the TF-IDF table labelled Top Specific Words, the word "paladino" was used 2,857 times in October of 2010. The word is connected to the hateful remarks that New York Republican gubernatorial candidate, Carl Paladino made. He related homosexuality to

	words	tf	month	idf	tfidf
1046	ýçùþëüåùéåæòîãüßîôêøâò	63	2010_12	2.484907	156.549119
239	obituary	89	2010_04	1.098612	97.776494
951	dmv	65	2010_10	0.875469	56.905468
716	gazette	30	2010_04	1.791759	53.752784
1070	ketchup	29	2010_08	1.791759	51.961025
573	ritchie	37	2010_04	1.386294	51.292891
622	gloria	34	2010_04	1.386294	47.134008
967	adsense	33	2010_08	1.386294	45.747714
2217	qatar	24	2010_12	1.791759	43.002227
2771	stylecolor	16	2010_10	2.484907	39.758506
2821	fontweight	16	2010_10	2.484907	39.758506
2842	990000	16	2010_10	2.484907	39.758506
2949	handtohand	16	2010_12	2.484907	39.758506
1182	harvard	55	2010_12	0.693147	38.123095
3137	kyke	15	2010_12	2.484907	37.273600
2854	paladino	15	2010_10	2.484907	37.273600
3071	lebanese	15	2010_11	2.484907	37.273600

Figure 3: Top Specific Words

pedophilia and insinuated that anyone who is a part of the LGBTQ+ community is a pervert [4]. The statement was extremely publicized and unfortunately many Americans agreed with Paladino's ideals. According to a Gallop poll, 40% of Americans thought that gay or lesbian relations between consenting adults should be illegal in 2010.[5] In other words, the plethora of

---

[4] James, Frank. "Carl Paladino: Gay 'Option' Not 'As Valid' As Straight" *NPR*, 2010.
[5] Gay and Lesbian Rights. *Gallup Poll*, 2010.

individuals who were vehemently against gay rights helped the advancement of negative speech during this time. The frequency of "paladino" demonstrates that the dialogue surrounding the LGBTQ+ community was faced with extreme hate and negative language.

Figure 3 demonstrates that in 2015, the word "Leelah" had a wordcount of 447 and was specifically important because the frequency prevailed for several months. A majority of the words that we analyzed had a significant rise in wordcount that lasted one month, but "Leelah"
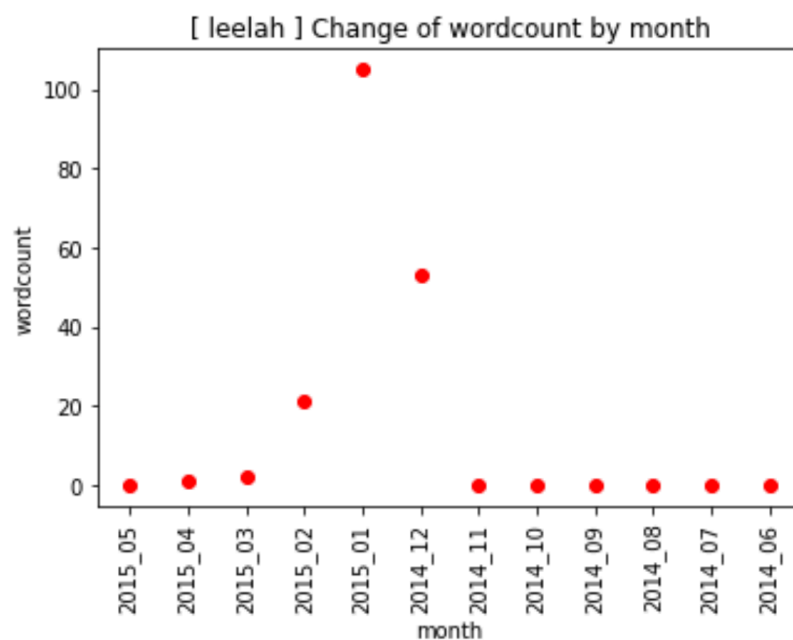
Figure 4: Leelah Wordcount Over 12 Months

rose frequently over the course of three months. We felt that the persistence of this word over time was important to analyze, as it indicated that it was connected to a special event. The word began circulating after the suicide of a young transgender girl named Leelah Alcorn that gained international attention because she posted her suicide note online.[6] Her death didn't just cultivate dialogue but also influenced law makers to propose a bill in Cleveland that banned gay

---

[6] Newman, Sarah. "Hang On, It Does Get Better: On Leelah Alcorn's Suicide Note" *PsychCentral*.

conversion therapy.[7] This particular moment of time was clearly extremely important for many within the LGBTQ+ community and facilitated legal action. Even though the word "Leelah" has a tragic connection, our research indicates that there was a notable shift over time in the context behind the overall language being used.

In order to analyze the general language used regarding the LGBTQ+ community, we compiled the top grammatical adjective-noun pairs regarding gay men and lesbians. There was a notable difference in the language used between 2009 and 2015 in that over time, we found an increase in accepting language and a decrease in words with a negative connotation as portrayed in the graphs. The trend is seen within the pairings regarding both gay and lesbian, but is more prevalent within the words referencing gay.
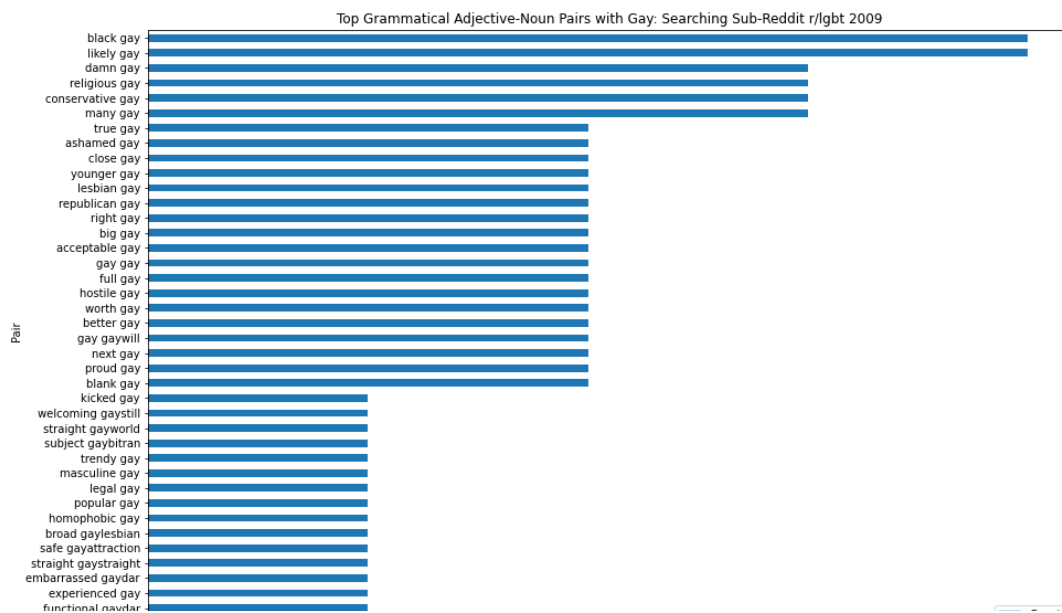

Figure 5: Gay Pairings 2009

[7] Capehart, Jonathan. "Obama comes out against 'conversion therapy' to support 'Leelah's Law'" *The Washington Post,* 2015.

As seen above in Figure 5, some of the pairs with the highest wordcount referencing gay in 2009 were "religious gay," "conservative gay," "damn gay," "acceptable gay," and "ashamed gay." The pairings of religious" and "conservative" were interesting because they reference the religious opposition to homosexuality. According to a Pew Research poll, 84% of individuals who categorize themselves as religious believed that homosexuality should be discouraged [8]. This considerable number indicates that religion and homosexuality is a point of contention for many individuals within the LGBTQ+ community. There is a strong relationship between conservatism and being against homosexuality. Although there are a handful of LGBTQ+ individuals who identify as conservatives, it is likely the grammatical pairing is meant to portray the hostility between the two groups. The word "ashamed gay" is likely in reference to the
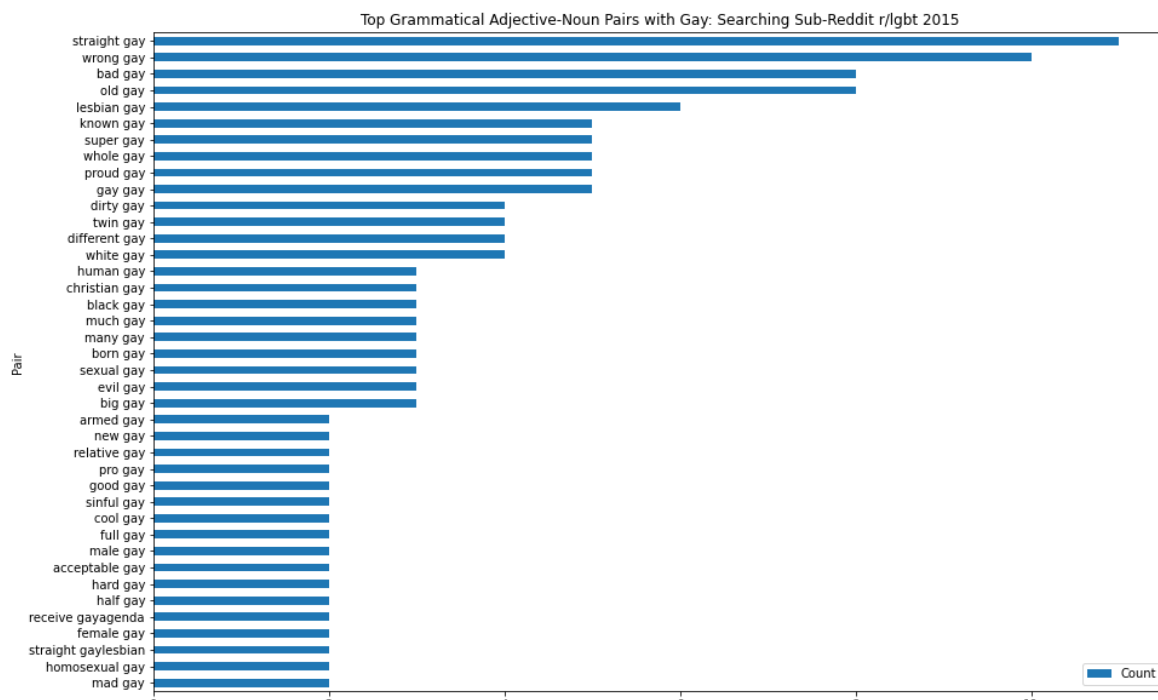

Figure 6: Gay Pairings 2015

---

[8] Views About Homosexuality. *Pew Research Center,* 2014.

countless individuals within the community who have been unable to accept themselves due to the heteronormative nature of society. The frequency of these words indicates that much of the language used regarding the LGBTQ+ community had a substantial negative connotation in 2009. Although the 2015 data still contains many negative word pairings, there is an obvious decrease. As seen above in Figure 6, in 2009 the count for the word pairing "proud gay" was at 2.0 but increases significantly to 5.0 in 2015. Another interesting pairing was "human gay," as this indicates that language was being utilized to articulate that members of the LGBTQ+ community are people and not perverse outcasts.

Similarly, the pairings found with lesbian portrayed in Figure 7 indicate a similar trend. In 2009, some of the highest and most compelling pairings were "blackballed lesbian" and "repulsed lesbian."
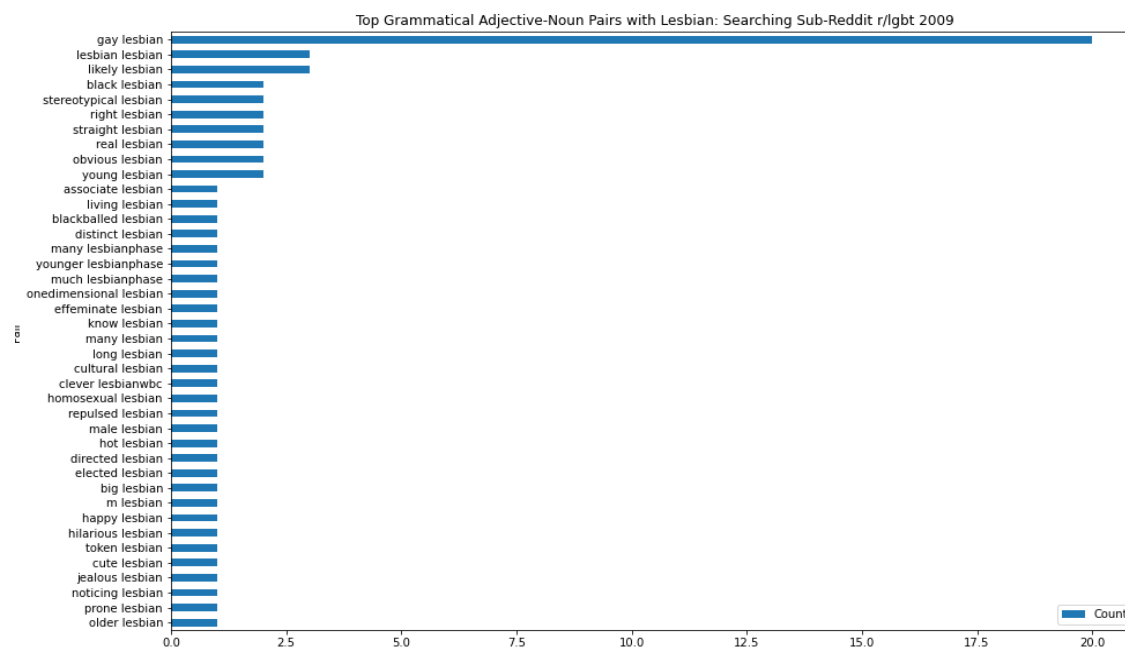

Figure 7: Lesbian Pairings 2009

The term "blackballed lesbian" is used to describe someone who is rejected or outcasted due to their sexuality. The frequency of this term highlights that many members of the LGBTQ+ community were faced with rejection and disownment. Furthermore, the pairing "repulsed lesbian" also portrays a tone of disgust towards the community. Interestingly, the word "phase" was placed directly behind our pairing of lesbian a handful of times in 2009, which is a highly controversial term to describe an individual's sexuality. As seen in Figure 8, by 2015 many of the adverse comments that were seen in 2009 disappeared. The 2015 pairings still contain phrases with negative connotation, like "fat lesbian" and "wrong lesbian," but are significantly
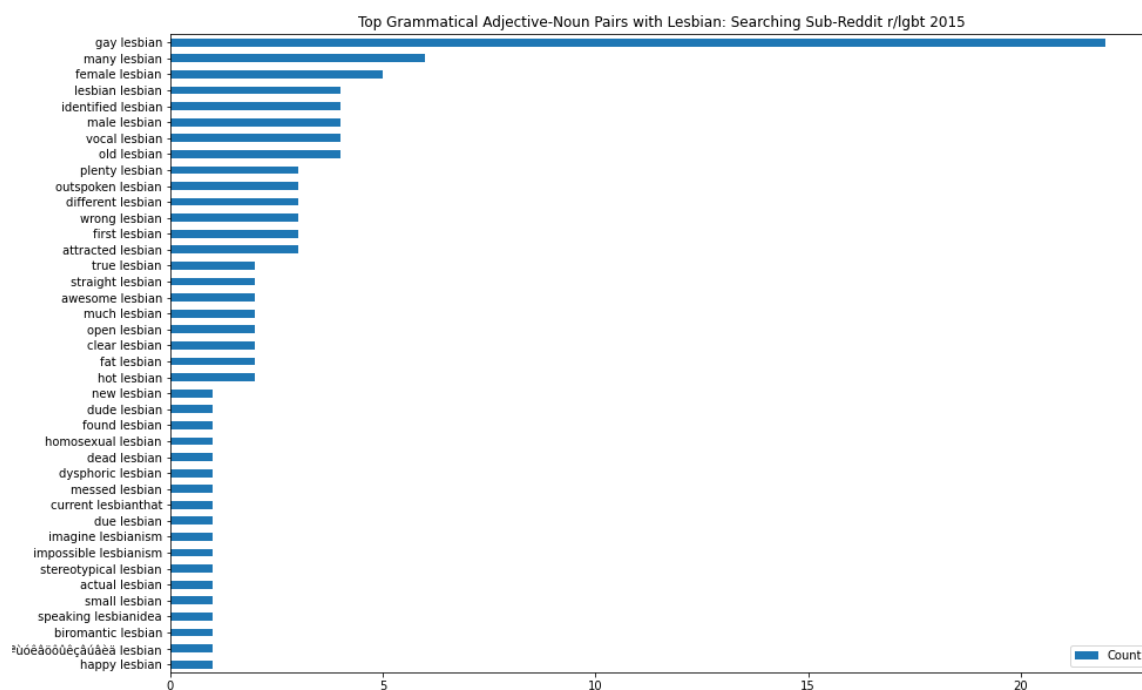

Figure 8: Lesbian Pairings 2015

less harmful in connection to the community. Words like "repulsed," "phase," and "blackballed" are distinctly utilized to damage individuals who belong the community. The trend that is

portrayed through these visualizations clearly indicate that the language used to describe the LGBTQ+ community has shifted towards a positive overtone.

Conclusion

Throughout our research, we discovered that the language used regarding the LGBTQ+ community has become more positive over time. Changes in legislation and current events largely influence the dialogue and have been the catalyst for the change in language. In addition, we identified that many words that experienced a spike in wordcount were deeply personal to the LGBTQ+ community. Although research has been conducted previously regarding the attitudes of individuals towards the LGBTQ+ community, researching the specific language is a unique perspective. Understanding the language shift is critical in analyzing the broad scope of LGBTQ+ acceptance.

Work Cited

Burelli, David. "*"Don't Ask, Don't Tell"*: The Law and Military Policy on Same-Sex Behavior."

  Congressional Research Service, 2010.

Capehart, Jonathan. "Obama comes out against 'conversion therapy' to support 'Leelah's Law'"

  The Washington Post, 2015.

Department of Defense. "2011 Demographics, Profile of the Military Community," 2011

Gay and Lesbian Rights*. Gallup Poll*, 2010.

James, Frank. "Carl Paladino: Gay 'Option' Not 'As Valid' As Straight" *NPR*, 2010.

Newman, Sarah. "Hang On, It Does Get Better: On Leelah Alcorn's Suicide Note" PsychCentral.

Smith, Tom W., Jaesok Son, and Jibum Kim. "Public Attitudes Toward Homosexuality and Gay

  Rights Across Time and Countries." 2014.

Views About Homosexuality. *Pew Research Center,* 2014.

https://fas.org/sgp/crs/misc/R40782.pdf
(don't ask don't tell)

Burelli, David. *"Don't Ask, Don't Tell": The Law and Military Policy on Same-Sex Behavior.* Congressional Research Service, 2010.

https://download.militaryonesource.mil/12038/MOS/Reports/2011_Demographics_Report.pdf
(military stats)
Department of Defense. "2011 Demographics, Profile of the Military Community"

https://www.washingtonpost.com/blogs/post-partisan/wp/2015/04/10/obama-comes-out-against-conversion-therapy-to-support-leelahs-law/
(Leelah)
Capehart, Jonathan. "Obama comes out against 'conversion therapy' to support 'Leelah's Law'" *The Washington Post,* 2015.

Newman, Sarah. "Hang On, It Does Get Better: On Leelah Alcorn's Suicide Note" PsychCentral.

(Leela)

https://www.npr.org/sections/itsallpolitics/2010/10/11/130487561/carl-paladino-gay-life-not-as-valid-as-straight
(Paladino)
James, Frank. "Carl Paladino: Gay 'Option' Not 'As Valid' As Straight" NPR, 2010.

Smith, Tom W., Jaesok Son, and Jibum Kim. "Public attitudes toward homosexuality and gay rights across time and countries." (2014).

https://news.gallup.com/poll/1651/gay-lesbian-rights.aspx
Gay and Lesbian Rights. Gallup Poll, 2010.

https://www.pewforum.org/religious-landscape-study/views-about-homosexuality/
Views About Homosexuality. *Pew Research Center,* 2014.