**1.Fix of Introduction**

Nowadays, the context individuals consume on media has become shorter and provocative. When Twitter first came out, it was questioning if it is possible to convey one’s idea within 140 words. Contrary to the concern, the average length of tweets are only 28 words meaning the 140 words is more than surplus. This shows how brief the contents are formed. The usage of words has turned into simpler and provocative while discourse society is facing has become more diverse and dynamic than ever before. This project shows another aspect of this trend through pop music particularly in hip hop songs. Hip hop songs are relatively more dependent on the lyrics and its context having less dependency on its rhythm and melody. Therefore, by looking into the changes in lyrics of hip hop songs over time would give a hint for how people’s consumption patterns have changed.

This project specifically focuses on two linguistic characteristics in songs. First is how unique words are used in each song. This shows how diversely words are adopted to deliver the subject. More unique words used in a song means more diversity and more complexity. The approach to measure the level of diversity in vocabulary is achieved by counting the unique instances of words: for example, if a certain word appears more than twice in a song, only the first instance is kept leaving the rest duplicates out of the word pool.

Second is how broad audiences’ vocabulary would have to be. This focuses on how various subjects are dealt with in songs as well as how different context of matters are used in each song. As more different words are used in different songs, it is assumed that different matter is conveyed by distinctive voices not by the generic manner. In addition to the quantitative analysis of absolute count of words, proportion analysis is also included, which means how much percentile is covered by having certain coverage or word. For example, by covering top 10% of vocabulary, up to how much percentage of vocabulary is covered in a certain decade.

To analyze trends over time, this project covers hip hop songs from the first advent of it (1980) to the songs of today(up to 2020). The user is able to compare the changes by navigating dispersion pattern visually by decade and also to see what is the general pattern of vocabulary by seeing the generative art form of word dispersion.

Since this project shows an aspect of pop culture linguistically, the vinyl player interface is adopted as a visual inspiration. The user can explore the processed information as if one plays with analogue music player.

**2. TREATMENT**

1) overview of changes and trends

**How hip hop songs were born**

Before talking about the history of raps. It is necessary to define what is hip hop or what is rap. These two seemingly interchangeable concepts are not actually replaceable. In a word, hip hop is a cultural territory which encompasses a culture emerged from black communities in the united states. It covers DJing, BBoying, and of course raps. Even if this culture is no longer limited to a certain ethnic group now, it is how this cultural area was firstly born to this world. Rap, on the other hand, is the genre of music is a form of art using verbal gymnastics with party beats. That is how this genre firstly came out to this world. This black popular music was emerged in South Bronx during seventies. This was a cultural niche movement in that regional base and then spread to the entire nation manifesting ‘make voices’ for the suppressed people.

Not long after it takes one of the main stream of pop culture and now it is regarded as a representative genre heavily based on lyrics.

Shortly speaking, rap is a type of music created within hip hop culture represents current pop culture for the suppressed.

Considering that this music makes the most use of verbal elements with help of beat, this is a good source of analyzing how vocabulary usage pattern has changed over time and how vocabulary is adopted in pup culture. There is no wonder if other types of pop songs are more relevant to this analytical approach. According to the recent research conducted by “MusiXMatch’, which is mega data base of music and its lyrics, rap is the genre adopting the largest size of vocabulary with the average size of 1963 followed by heavy metal whose average is 1553.

**Change of ways to define problems**

Society has gone through many changes. When society changes, what is problem and how the problem became a problem and how it should be dealt with also change. Individuals used to see a problem as a problem following an objective standard, so called ‘objectivist’ perspective. This paradigm rolled toward a new point of view.

In this newer aspect, problems are not what a group defines as a problem to be treated. Rather the problems fell into a very individual compass where one could subjectively define it and put a filter based on one’s experience. No longer people contend to what has been discussed. They rather say things based on what they want to talk about.

This projects particularly also casts a light how the vocabulary changes have been made from the sociological angle by providing changes in social paradigm. Social discourse paradigm has changed from objectivist perspective to constructionist perspective. It means the validity of problem has shifted from genuine definition of the problematic phenomenon to the more subjective and individual interpretation of the object. These chronical changes are going to be covered in the later section.

**Influence from music to listeners**

The influence is not one-directional. Music also influences the listeners. According to the study, the contents can also shape the listeners’ behavior and consequently change the thought of the listeners too. Applying social cognitive theory (SCT; Bandura, 1994) can explain this influence. As kid learns the norm of society, the adults also learns what is regarded as a norm of the society through diverse media. Music, without doubt, is one of the most influential channel. SCT asserts that behaviors are often learned by watching others (Bandura, 2002). Audience members are more likely to emulate behaviors that are rewarded. For example, audience members who perceive characters to be similar to themselves pay more attention to those characters’ behaviors. Since hip hop gained its position as a prevailing culture pattern and agreed mainstream of the popular culture, it was no longer limited to a certain circle. Rather, more and more people could accept it and take it naturally. Therefore, we can say that lyrics and society are changing and affecting each other to a certain direction.

**Trend of vocabulary - Repetition**

If so, what is the direction that lyrics has been leaning toward? There is a research about how lyrical contents has been changing over past 25 years. And what drove the change?

According to the paper (The power of repetition: repetitive lyrics in a song increase processing

fluency and drive market success), there is a strong influencing variables. The repetitions have increased cause the audiences market demands more repetitions which eventually led songs to adopt more repetitions to get accepted more by the audiences. It brings up a foundational question that what makes people to prefer over the variations in a song? The answer is ease of processing. One benefit derived from having encountered a stimulus previously, or repetitive priming, is an ease of processing referred to as processing fluency. Fluency effects have been documented widely for language, including most often for individual words (Hutchins & Palmer, 2011). An ancillary benefit resulting from processing fluency, notably for aesthetic goods, is that the experience is typically more pleasant (Reber, Schwarz, & Winkielman, 2004). Consequently, we expect songs that are lyrically more repetitive (for instance, by repeating the chorus more often),and thus more fluent, to be generally preferred and adopted more quickly and broadly in the marketplace. As society changes, the influx that individuals have to deal with is getting greater, which leads one could feel fatigue to register the information. Accordingly, the easier process of input should be preferred over what requires more processing. Based on this, this project assumes that less diversity would be witnessed once the lyrics data is processed. The methodology to prove this premise will be discussed in the later section.

**Subject of songs**

Has subject of songs become diverse as the society discourse become diverse? Answer is that It has not been just more diverse than before, there has been a significant mainstream shift in topics and theme that songs are singing about. From the study (Peter G. 2019) which looks into this change from 1960 to 2010, the dominant contents of popular songs has gone through drastic changes and leaned toward so called ‘censorable’ context. This study particularly focuses on a specific content domain, such as sexuality, substance use, violence, or suicide.

This focuses on studies that do one or more of three things: (a) examine samples of top-ranked songs (as opposed to collections of songs that happen to contain a reference to a narrow issue); (b) deal with thematic categories that have played a central part in popular music and in the research literature on it (for instance, romance, sex, substance use, politics, violence); and/or (c) provide evidence related to trends over time. During 1980’s, so called “boy-girl-issue” took the majority of song themes. Christenson and Roberts (1998) analyzed the top-40 Billboard hits from 1980 to 1990 and found that, of 240 songs examined, 73% had a “love relationship” as the major theme. Even a casual glance at the current charts confirms the importance of boy–girl relationships on the popular music agenda.

References to sex have also become more explicit over time. Primack, Gold and colleagues (2008) examined a sample of 279 top songs from 2005 pulled from several different charts (Pop, Hot-100, Country, R&B/Hip-hop, Rap, Mainstream Rock, and Modern Rock) for the frequency and tone of sexual references. Some 37% of songs made some sort of reference to sexual intercourse. Sexualization increased markedly from 1959 to 2009.Roberts,Henriksen, and Christenson (1999) examined 1,000 popular songs from the years 1996 to 1997. Of this sample, 18% of songs contained references to illicit drugs, 17% to alcohol, and 27% referred to either or both. Overall, 33% contained a substance reference of some kind. By seeing this trend changes, we can say that the context of popular music has headed toward more provocative subject matters. This of course affects the vocabulary and words used for the songs.

**Word mutation due to censorship**

If the previous trend has general tendencies only, the variations of diversity in vocabulary cannot be fully explained. The other critical variables that brought the wide variety of slangs, which is emerge of censorship in popular music. This hidden factor pushes the lyrics to take divergent word usage to express the same idea. From 1998, Law applied stronger censorship toward music to regulate and discipline popular culture. It creates criteria and boundaries by which popular culture are judged. Steve Redhead describes as ‘the fertile deconstructed terrain where legal theory, deviance and cultural studies collide’. Because much popular music has been targeted at a youthful audience, the sexual and violent content in lyrics has been the focus of conservative campaigns to ban, censor or prepress. To avoid this censorship, music lyrics had to adopt sideways, which eventually led to using metaphors related to the violent context. The table referenced from (You Know What It Is: Learning Words through Listening to Hip-Hop,2011) shows the snap shot of the trends. Most of word are deeply related to the violent context which could be possibly censored by the strict regulation, even if the word itself would not be filtered out by the censorship.

Table

Description automatically generated

These strategy accelerated emergence of diversity in words from using slangs.

2) Analysis Methodology

**Method to verify the repetitiveness within songs**

The premise of repetitiveness was made based on the assumption that the listeners market is more fond of repetitive lyrics over diverse usage of words due to the mind fatigue to process the information. This assumption is verified by counting unique words of hip hop songs from 80’s to 2010. The technical approaches to this verification step is made by following steps.

1.Collecting step-selection of artists

To see the representative trends of songs, notable hip hop artists were chosen. If the sample artists ranges to the niche like independent artists, it is hard to say that it represents mainstream or general trend of hip hop. Therefore, popular and generic open source was chosen which is Wikipedia. Searching by ‘notable hip hop artists’, over 200 hundred artists group is chosen including solo artist to hip hop groups.

2. Collecting step -selection of songs of each artist

After collecting all the songs from each artist, two problems occurred. First problem is that there are too many songs for each artist. Sometimes the songs only consist of one or two words or none. Including all of these data could be misleading because some songs are experimental songs that actually are rarely listened to. To avoid this misconception, targeting songs to be collected becomes selective. From the data source of this project, which is Genius API, the top 10 most frequently fetched songs are chosen to follow the popularity and generality of songs of the notable artists.

3.Processing step

After collecting the lyrical data from each song, unique words are picked leaving duplicates of the word from the song. The unique word count indicates two possible aspects of each song. If a song is absolutely shorter than the others, even if the song is less repetitive, the unique word count could me smaller than the more repetitive but lengthy song unless the longer song uses extreme level of repetitiveness. The other aspect is that the song with less unique word count is actually more repetitive than the others with greater unique words count. The latter case is what this project is seeking toward though, the extreme case risk exists still. To eliminate this loophole and some other outliers from the data, top and bottom 20% of samples are going to be opted out in the later step.

4.Visualizing

To show the gradual changes, fragmentation of timespan is not going to be applied in this visualization. The x-axis of the visualization will be the timespan and the y-axis is for the number of unique words in each song. The range of the value is dramatic. For example, for the repetitive and short songs, the count is barely over hundred or even sometimes smaller than a hundred while lengthy less repetitive songs take more than 2,500 counts. The first part is for showing the overall distribution of the count . Therefore, the less selective data is going to be presented including outliers. Then the filtering process is going to be done by removing the outliers which has extremely great word count as well as extremely small cases. After removing these two polar values, the general tendency can be acquired as well as removing outliers.

**Visualization of methodology**

Even if the unique word count approach is fairly intuitive, the titling could be misleading if the unique indicates unique cases across the songs or genre etc. To get rid of this confusion, this methodology specifically needs to be visualized. To show it clearly, instead of pulling the entire lyrics of one song, the partial lyrics is chosen and presented. Each words are highlighted with color and the duplicates are muted and removed from the process. The word count for the word changes dynamically to convey this idea to the audience.

**Method to show the trendy words of each decade**

Unlike the overall distribution of word count of hip hop songs, the trend change is based on certain size of timespan. In this study, a decade range was chosen for the span to see the trend changes. After examining the data set of lyrics, top 10~20 words are mainly common linguistic elements which are such as pronoun, articles or prepositions. This actually cannot fully represent the trend changes over time. So these common elements in the sentences are to be removed. After removing common elements, still there are some other words left relatively having more significance but not trendy nor representational. However, these words are going to be kept not to be subjective and not to harm the originality of the data. Then another problem comes up which is similarity among the decades. Even if the lyrical context has gone through dramatic changes over time, the variation of change has a limit of variation. It take more than several decades to change the vocabulary pool entirely that it is hard to see the dramatic changes in basic vocabulary in songs accordingly. To see the changes in vocabulary distinctively, it is inevitable to broaden the word pool to pay attention to. Therefore, this project decides to set the range of 100 most frequent words. Unfortunately, it is visually overwhelming placing the hundred words and compare the changes one by one that it utilizes the visual language that this project is using. That is analogue audio player interface, specifically equalizer bars. The words will be played as if it plays a song which consists of the 100 frequent words and let the words flow to see the general pattern of usage of the words.

The symbolic artists in the era will be displayed as a background to create a mood of the era. Then user explore each decades by opening representing albums of each decade and play the song made of 100 frequent words of each era.

**Visualization of individual words**

Now the audience gets the sense of word usages in each era, and within songs. It is phase to see the position of individual words within a whole distribution. It could represents the distribution pattern of the word in each decade. It is visually overwhelming to show the granularity that relative and comparative distribution would be the feasible approach. Considering the performance on client side processing, the level of interactivity needs to be carefully chosen. The user exploration is limited to two user interactions to explore the details of this analysis.

First user experience is to see the relative amplitude among the entire vocabulary used. To achieve this, first the data pool has to be re-organized. Before this stage, all the lyrical data is modularized by songs. Therefore, another data format will contain a data pair which are as below.

|  |  |
| --- | --- |
| Vocabulary | Frequency |
| “xxx” | “###” |

By reading this "word & frequency” pair within in the timespan. The frequencies of each vocabulary can be quantified in the visualization.

The how can these word and frequency distribution be visualized keeping the visual consistency? This another problem because the number of songs selected are not consistent. The number of songs during early stage is less than the number of samples selected for the later times like 2010 because the selection has to be from the notable artist pool and top 10 frequent songs of each artists. Inevitably the number of song to represent each era varies by time. Therefore, the proportional approach has to be taken. More specifically speaking with an example, if there are ten songs for the 80’s because the number of songs produced during 80’s cannot be as large as the songs made during 10’s due to production infra differences and market size differences. Therefore, the number differences should not be visualized in this visualization. This visualization would rather focus on the proportional differences across the decades studied.

several chapters

The treatment builds on the introduction:

initial premises and the questions you promised to explore!

Talk about history

**analysis/critical discussion** of this material.

You should both refer to authorities with whom you agree and then argue against those with whom you disagree!

References

Easy listening? An analysis of infidelity in top pop, hip-hop, and country song lyrics over 25 years Cassandra Alexopoulos1 and Laramie D. Taylor