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 280 words. Contrary to the concern, the average length of tweets are only 28 characters meaning the 280 words is more than surplus. This partially reflects the current trends of communication among people. 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 comparably 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 the contents consumption pattern of people.

This project specifically focuses on two parts of linguistic characteristics in songs. First is how unique words are used in each song. This shows how diversly words are adopted to deliver the subject. The more the unique words are used in a song, it means more diversity is used in a song and more complexity is adopted. The approach to measure the level of diversity in vocabulary is achieved by leaving 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 the dispersion across the songs. 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 through the songs or at least different approaches are adopted to convey the same idea as others. In addition to the quantitative analysis of absolute count of words, proportion-wise 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 the shift of trends overtime, hip hop songs from the first advent of it (1980) to the songs of today(up to 2020) is covered. The User is expected 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. User would explore the processed information as if one plays with analogue music player.

Data sources are from genius API (www.genius.com) , the largest lyrics archives and are processed by node.js and turn into simplified json format.