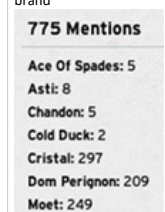


Summary - Previous Studies of Rap and Music Lyrics

Tuesday, January 5, 2016 10:24 AM

| No | Authors | Reference | Year | Topic | Genre of Songs | Data | Size | Analysis | Special Visualizations | Key Variables |
|----|---|---|------|---|--------------------------|---|---|---|--|---|
| 1 | Daniels, M. http://poly-graph.co/shorty.html | The Etymology of Shorty in Hip Hop: An Absurdly Nerdy Look at How Hip Hop Invented the Most Important Slang of Our Time | 2015 | Analysis of the use of the word Shorty in rap lyrics as well other associated words over time | Rap/HH | Lyrics from RapGenius.com and album release data from MusicBrainz | NA | Counts of the usage of the word <i>shorty</i> , <i>shawty</i> by year and by meaning | Stacked area charts for various meanings of Shorty by Year as well as bar charts of other word usage | Single Words shorty |
| 2 | Daniels, M. http://poly-graph.co/timeless/ | The Most Timeless Songs of All-Time: Using Spotify to Measure the Popularity of Older Music | 2015 | Analysis of play counts of various tracks and the associated artist by year of release | Variety | Spotify data cover full range of years from 50s until 2015 | NA | Frequency analysis of Spotify play counts by various artists and tracks by year | Interactive rank charts, bar charts and time lines | Popularity of tracks via Spotify |
| 3 | Daniels, M. and Beacham, K. http://poly-graph.co/labels/ | <i>The Most Successful Labels in Hip Hop</i> | 2015 | Ranking and Analysis of Success of Record Labels | Rap/HH | No of weeks that their artists tracks were on Billboard's Hot Rap Artists 1989 to 2014 | 16 years by 52 weeks of top 25 on bill board | Time and geographically based statistical analysis of the performance of the top rap record lables | Interactive Table and Sprite Charts by Label, Artist and Year; Interactive US Map of Label performance by year, Artist or Label | No weeks on Bill Board Hot Rap |
| 4 | Malmi, E. http://mining4meaning.com/2015/02/13/raplyzer/ | Algorithm That Counts Rap Rhymes and Scouts Mad Lines | 2015 | Presents Raplyzer, a computer program which automatically detects rhymes from rap lyrics and which is used to rank popular rappers based on their average Rhyme factor. Presents another program called BattleBot, which is a search engine for rhyming rap lines based on the algorithm used in Raplyzer. | Rap/HH | Scraped the lyrics of 94 rap artists from a lyrics website. | 10,082 songs. | For each artist, I computed the Rhyme factor averaged over all the songs of the artist (based on assonance rhyming). Rank ordered and then compared rhyme factor score by number of unique words (ala Daniels) | Scattergram of rhyme factor by unique words for each artist. | Rhyme factor for assonance |
| 5 | Malmi, E. http://arxiv.org/pdf/1505.04771v1.pdf | Machine-Learning Algorithm Mines Rap Lyrics, Then Writes Its Own | 2015 | The highly structured nature of rap makes it particularly amenable to computer analysis. And that raises an interesting question: if computers can analyze rap lyrics, can they also generate them? | Rap/H | A list of 104 popular English-speaking rap artists and scraped all their songs available on a popular lyrics website. | In total, we have 583 669 lines from 10 980 songs | Our approach is based on two machine learning techniques: the RankSVM algorithm, and a deep neural network model with a novel structure. For the problem of distinguishing the real next line from a randomly selected one, we achieve an 82 % accuracy. We employ the resulting prediction method for creating new rap lyrics by combining lines from existing songs. In terms of quantitative rhyme density, the produced lyrics outperform best human rappers by 21 %. | None | Rhyme factor for assonance |
| 6 | Mauch, M. et al. http://rsos.royalsocietypublishing.org/content/2/5/150081 | The Evolution of Popular Music: USA 1960–2010 | 2015 | Using music information retrieval and text-mining tools, analyze the musical properties of approximately 17000 recordings that appeared in the charts and demonstrate quantitative trends in their harmonic and timbre properties. Use these properties to produce an audio-based classification of musical styles and study the evolution of musical diversity and disparity, testing, and rejecting, several classical theories of cultural change. Investigate whether pop musical evolution has been gradual or punctuated. Show that, although pop music has evolved continuously, it did so with particular rapidity during three stylistic 'revolutions' around 1964, 1983 and 1991. Conclude by discussing how our study points the way to a quantitative science of cultural change. | Variety including Rap/HH | Focused on songs that appeared in the US Billboard Hot 100 between 1960 and 2010. We obtained 30-s-long segments of 17094 songs covering 86% of the Hot 100. The earlier years had more missing songs than the later years. | 30-s audio of 17094 songs | LDA was used to determine the underlying harmonic and timbral topics - 8 in all. Each track was represented as a distribution over eight harmonic topics that capture classes of chord changes (e.g. 'dominant-seventh chord changes') and eight timbral topics (T-topics) that capture particular timbres (e.g. 'drums, aggressive, percussive', 'female voice, melodic, vocal', derived from the expert annotations), with topic proportions q. These topic frequencies were the basis of the analyses. Topics then used to constructed a taxonomy of 13 styles by k-means clustering on principal components derived from our topic frequencies. This was the base of the analysis of the rise and fall of the styles as well as the changing audio features. | Single Chart showing time series graphs for each of the 8 <i>harmonic</i> and 8 <i>timbre</i> audio feature topics. Specialized combo chart showing dendrogram of styles/genres and a spindle graph of frequencies of those styles by years. Combined Vertical Time Series/Boxplot of 4 Diversity Measure by Time. Quarterly pairwise distance matrix of all the songs in the Hot 100 based on 16 audio topics. Graph of Foote Novelty kernel for each year. | Audio features of songs - harmonic and timbre |
| 7 | Powell-Morse, A. | Lyric Intelligence in | 2015 | Using song lyrics, tries to answer questions about which genres | Variety including | Analyzed 225 tracks from | 255 tracks | Analysis included overall scores by year, artist sex and year, genre and year, and avg. word count by | Time line charts and bar charts | Lyric Intelligence |

| | | | | | | | | | | |
|----|--|--|------|--|--------------------------|---|---|---|---|--|
| | http://seatsmart.com/blog/lyric-intelligence/ | Popular Music: A Ten Year Analysis | | and artists have the most and least intelligent lyrics (based on readability scores). | Rap/HH | Billboard charts for Pop, Country, Rock, and R&B/Hip-Hop for any given year from 2005-2014. | | year; average score by artist over the years by genre; 10 dumbest and 10 smartest songs over the time span. | | |
| 8 | Ruiz, E. http://poly-graph.co/camron/ | <i>Why Cam'ron's Hey Ma is Peak Nostalgia</i> | 2015 | Examines 150 of the top songs from 2001-2005 and culled them down to sixteen tracks looking at their current | Rap/HH | Billboard ranking and Spotify ranking | 16 tracks | Relationship btw. Billboard percentile at year of release and Spotify ranking in 2014 | Parallel coordinate chart displayed horizontal | Popularity of tracks via Spotify |
| 9 | Thompson, J. http://harvardolitics.com/covers/evolution-rap/ | The Evolution of Pop Lyrics and a Tale of Two LDA's | 2015 | Inspired by The Evolution of Popular Music: USA 1960–2010 (#6 above) which utilizes audio signaling processes to understand the evolution, this paper explores investigates the possibility of tracing the same evolution through lyrics. | Variety including Rap/HH | Used artists and songs from Evolution of Music study (songs that appeared in the US Billboard Hot 100 between 1960 and 2010.). Lyrics obtained by ChartLyrics API | ~80% of 17094 songs (=13672) | Used Latent Dirichlet Analysis (LDA) to determine underlying topics. After exploration settled on 25. Unfortunately these did not do well in predicting the 13 genres assigned in the original evolution study (only 16% predicted but rap once one of the better ones-60%). | Word Clouds, Topic Chart, Heat Map (all created with D3js). | Analytically derived Topics of popular music |
| 10 | Big Data, MC https://medium.com/cuepoint/riding-dirty-the-science-of-cars-and-rap-lyrics-21b8404a9c4d#.z2i9yphwo | Riding Dirty: The Science of Cars and Rap Lyrics | 2014 | Which car makes appeared most frequently in rap lyrics from 1996-2014 | Rap/HH | Rapgenius.com lyrics (no mention of actual sample) | Measured the frequency of various makes across the years. | Frequency of mention in lyrics by car make and across time | Time series line charts | Car Brands |
| 11 | Corbett, J. https://www.rit.edu/cla/communication/sites/rit.edu/cla/communication/files/images/2141%20-%20Corbett%20Jarisa%20-%20Senior%20Thesis.pdf | Flexing, Twerking, and Popping Mollies: A Textual Analysis of Hip-Hop Songs and Music Videos | 2014 | Textual analysis of hip-hop songs—both lyrics and music videos. Using standpoint theory, the study examines hip-hop songs and music videos' gender dynamics, the power structure between men and women, and popular themes. | Rap/HH | A list of 25 hip-hop music videos was compiled, based on the top downloaded hip-hop music videos on iTunes as of September 2014. | 25 of top downloaded hip hop music videos from itunes in 2014 | A codebook was created whose purpose was to identify coding categories for analyzing the lyrics and music. For lyrics, these were the subjects analyzed: the occurrence of derogatory words, references alluding to money and extravagant living, references to having or getting any number of women and descriptions of their appearances, mention of drug and alcohol use on the performers themselves, the use of drugs and alcohol on women, the mention of partying, descriptions of sexual activities, shows of masculinity through money, sexual prowess or acts of violence. For music videos, these were the subjects analyzed: the gender of the hip-hop artist(s), hip-hop artists' style of dress, the presence of dancers, dancers' style of dress (if applicable), dancers' gender(s), the style of dance choreography, the use of a 'video vixen', the overall theme of the music video. Frequency analysis was done on list for the various subjects. | None | Sexual relations and prowess, misogyny, drug and alcohol use, partying, money and extravagant living |
| 12 | Daniels, http://poly-graph.co/outkast.html | <i>Outcasts in Charts: Look at Outkast's History Through the Lens of Data</i> | 2014 | Analysis of Outcasts' use of various words and themes | Rap/HH | Data based on lyrics and themes in various tracks over the years | Specific Word and theme usage from 1985 to 2012 | Time line and bar charts used to analyze use of specific words and themes | Stacked area charts and bar charts | Specific words - skeet, crunk, playa, guns, drugs, the south, pimpin, cars/lacs |
| 13 | Daniels, M. http://poly-graph.co/vocabulary.html | <i>The Largest Vocabulary in Hip Hop: Rappers, Ranked by the Number of Unique Words Used in their Lyric</i> | 2014 | Analysis of Lyrics by number of unique words in track | Rap/HH | Lyrics taken from Genius.com | 35K words from first albums of 85 prominent Rappers | Rank ordering of Rappers (vocabularies); special; | Interactive Ranking Ordering of Rappers by unique word counts by region | Lexical diversity |
| 14 | Hemphill, T. http://www.mappersdelight.net/ | Nas: Maximum Distance. Minimum Displacement. | 2014 | Used the Hip Hop Word Count's new semantic analysis results to extract all geographic mentions from the complete bodies of work of 12 rappers. These locations were translated into geo coordinates which were then made into points that plotted the robot arm's movements. The robot arm drew each path while holding a light pen. | Rap/HH | Tahir Hemphill's Hip-Hop Word Count database | Body of works of 12 artists from 50,000 rap songs from 3,000 artists, from 1979 to the present | Semantic analysis and extraction of geographical mentions. Locations translated into geo coordinates. | Light pen display of moving from point to point in the geographical mentions. | Geographic mentions |
| 15 | Hemphill, T. http://www.tahirhemphill.com/picasso-baby/ | <i>Picasso Baby!</i> | 2014 | Examines social relationships between rap artists and fine artists using mentions of the fine artists (e.g. Picasso) in the lyrics of the rap artist (say Jay-Z) | Rap/HH | Hip Hop Database including lyrics from 50K tracks and 3000 artists | 50,000 rap songs from 3,000 artists, from 1979 to the present | Creates bi-modal graph: artist creates track; artist name appears in track lyrics | Interactive SNA graph | Named (fine art) artists |
| 16 | Mussman, S., Moore, J. and Coventry, B. http://www.cs.purdue.edu/homes/moore269/docs/music.pdf | Using Machine Learning Principles to Understand Song Popularity: Can Lyrical Content Predict Song Trajectories on the Billboard 100? | 2014 | Use song lyrics to predict the ranking of a song on next week's Billboard's Hot 100. Does this by predicting song popularity trajectory labels and then utilizing the trajectory labels to predict song rankings. | Variety | MusiXmatch (MXM) bag-of-words version of the Million Songs Dataset (MSD) from 1999-2008. The Hot 100 songs by week were scraped from BB's Hot 100 site -- this is the set to predicted. | Given the year and the songs for which lyrics were available, the dataset was reduced to 940 songs. | The analysis was fairly complicated. Essentially, they gathered a series of NLP measures for each track - top 3 songs, average word length, overall memorability of songs based on rarity determined by tf-idf, song sentiment (polarity, pos-neg, subjectivity - along with a series of attributes from the MXM database including artists familiarity scores, artist hotness, year released, energy and tempo of song, and loudness of the song. These features were used to predict various trajectory labels (e.g. songs in the top {10,20,50} for a week or average change in rankings > {5,10}). These trajectory labels were then used to predict the next weeks ranking. | Standard bar charts, line charts, and scatterplots | Word frequencies, entropy, and sentiment. |
| 17 | Raschka, D. | NA Machine | 2014 | Building a classification model | Variety | Downloaded the | Hand-labeled a | Focused on optimizing precision and recall via | Stacked bar | Sad vs. happy |

| | | | | | | | | | | |
|----|--|---|------|---|--------------------------|---|--|--|---|------------------------------------|
| | http://sebastianraschka.com/Articles/2014_musicmood.html | Learning Approach to Classify Music by Mood Based on Song Lyrics | | that can automatically predict the mood of music based on song lyrics, i.e. classifying them as happy or sad.. | | lyrics from LyricsWiki, filtered out songs for which lyrics were not available, and automatically removed non-English songs using Python's Natural Language Toolkit. | subset of 1200 songs: 1000 songs for the training dataset and 200 songs for the validation dataset | the F1-score performance metric rather than optimizing the overall accuracy – I was primarily interested in filtering out sad songs. the differences between the different pre-processing steps and parameter choices were rather minor except for the choice of the n-gram sequence lengths. | chart of number and one for percentage of happy vs. sad for 10 periods since 1960-2010 | emotion |
| 18 | Sasaki, S. et al. https://staff.aist.go.jp/m.goto/PAPER/ISMIR2014sasaki.pdf | LyricsRadar: A Lyrics Retrieval System Based on Latent Topics of Lyrics | 2014 | Presents a novel system - LyricsRadar - "that enables users to interactively browse song lyrics by visualizing their "topics" rather than relying on simple word search. | Variety | Utilizes authors' database of popular Japanese (J-POP) songs and popular English songs taken from the Musical Lyrics Database. | 6902 J-POP songs (1847 artists, 2285 song writers, and 26K words) and 5351 English songs (487 artists and 36K words) | Employed Latent Dirichlet Allocation (LDA) to derive 5 topics and their weights for songs and artists. The weights were derived using key noun, verb, and adjective components from the lyrics (derived using "MeCab") after the elimination of stopwords and the weighting of the words via "idf." | Lyrics Radar Chart to both display and interactively search for songs and artists; 2D cluster display of songs and artists. Both are based on the 5 topic profile rather than single topic placement. | Topic profile for 5 derived topics |
| 19 | Ying, T., Doraisam, S. and Abdullah, N. http://scialert.net/fulltext/?doi=jas.2015.289.294&org=11 | Lyrics-Based Genre Classification Using Variant tf-idf Weighting Schemes | 2014 | Study an approach to lyrics based musical genre classification was presented which utilizing mood information. From the analysis of the lyrics text in the data collection, correlation of terms between genre and mood was observed. Based on this correlation of terms, new weighting equation with combine weights from genre and mood was introduced and implemented in two different ways. | Variety but not Rap/HH | Musical genre classification experiments were performed using a test collection consists of 1000 English songs. A total number of 10 genre types including pop, blue, country, folk, R and B, reggae, grunge, punk rock, soul and metal in test collection were selected. | 1000 English songs classified by genre. Mood of test set coded by 5 undergrads into happy, sad, angry, relaxed, calm, gloomy, romantic, confident, disgusted and aggressive. | Combined all documents of single mood (10) class and then computed specialized weighted tf-idf for 10 docs and looked at correlations | None | Genre based on mood |
| 20 | Cundiff, G. http://www.studentpulse.com/articles/792/4/the-influence-of-rap-and-hip-hop-music-an-analysis-on-audience-perceptions-of-misogynistic-lyrics | <i>The Influence of Rap and Hip-Hop Music: An Analysis on Audience Perceptions of Misogynistic Lyrics</i> | 2013 | Analyzed the misogynistic lyrical content of 20 popular rap and hip-hop songs found BB Hot 100 from 2000-2010, as well as their impact on audience views of violence against women | Rap/HH | Selected Billboard Hot 100 tracks on which to perform qualitative content analysis | 20 tracks | Content analysis determining the frequency of mentions of various types of misogyny (demeaning language, rape/ sexual assault, sexual conquest and/or physical violence) | Pie charts of % of types of misogynistic lyrics (out of all mentions) and % by "level of mentions" - low (1-2), med(3-4), high(5+) | Misogynistic lyrics by type |
| 21 | Higa, S. http://www.stephaniehiga.com/posts/analyzing-rap-lyrics-with-python.html | <i>Analyzing Rap Lyrics with Python</i> | 2013 | The most beloved car brand in hip-hop? | Rap/HH | Created list of car brands by manufacturer and country using manufacturer's list. Originally 10 countries and 2599 brands reduced to 4 countries – Germany, Japan, UK and US – and 153 brands. | Black Youth Group Rap Lyrics Database contains lyrics for all of Billboard Music's rap songs from 1989 through 2009. [Note DB no longer exists] | Frequency counts by brand and year. Note: provides Python code. | Google Charts | Car brands |
| 22 | Abraham, T., Koul, N. and Morales, J. http://people.ischool.berkeley.edu/~nikhitakoul/capstone/index.html | <i>R.A.P. - Rap Analysis Project</i> | 2013 | Applied machine learning techniques and data science principles to a database of rap lyrics from 1980 to 2015. After an active exploration of the data, focused on 'hit prediction', particularly on what it takes to make it onto the weekly Billboard Top 100 charts. | Rap/HH | Genius.com database of rap lyrics | 24,175 entries. This dataset contained 1,491 rap songs that had successfully made it onto the top 100 | General analysis of % of Billboard 100 rap songs by month and year. Investigated lyrics for brand names, geographical locations, crime & geography, and theme. Trained an support vector machine (SVM) on the features of song topic, vulgarity, and release date. Classified test dataset of 600 songs (300 successful and 300 unsuccessful) achieving over 71% accuracy for classifying the success of songs. Utilized AlchemyAPI for entity identification and Mallet for SVM. | CartoDB maps, relationship chart (% by mon and yr), stacked area charts. Relied on CartoDB, d3js and "Tableau" for charting. | Prediction of Hits on BB |
| 23 | Chaey, C. http://www.fastcompany.com/3007753/hip-hop-word-count-living-breathing-database-every-word-in-every-rap-song-ever | <i>Hip-Hop Word Count Is a Living, Breathing Database of Every Word In Every Rap Song Ever</i> | 2013 | Discusses the analysis mentions of brand products in Tahir Hemphill's Hip-Hop Word Count database which includes more than 50,000 rap songs from 3,000 artists, from 1979 to the present. T | Rap/HH | Tahir Hemphill's Hip-Hop Word Count database | 50,000 rap songs from 3,000 artists, from 1979 to the present | Focus on brand mentions of Champagne by brand  | Streamgraphs champagne brands over time; also US Map of mentions by artist/track location | Champagne brands |
| 24 | Sterckx, L. http://lib.ugent.be/fulltxt/RUG01/002/033/229/RUG01-02033229_2013_0001_AC.pdf | Topic Detection in a Million Songs | 2013 | Music Information Retrieval (MIR) is the interdisciplinary science addressing this potential, developing techniques including music recommendation. This work studies the use of themes in lyrics for this matter, using statistical analysis to detect topics. | Variety including Rap/HH | The main dataset used for this research is the so-called 'Million Song Dataset' (MSD) with metadata for 1,000,000 songs. This metadata is matched with 237,662 lyrics from commercial lyrics catalogue, | 237,662 lyrics from commercial lyrics catalogue, 'musiXmatch' and a dataset containing 8,598,630 social tag assignments, 'Last.fm'. | Labeled Latent Dirichlet Allocation, a super-used topic model, was trained using a labeled subset and was used for classification, which obtained results competitive with baseline performers but no large overall improvement. Latent Dirichlet Allocation, an unsupervised topic model was inferred from the corpus of lyrics, and evaluated according to semantic coherence and interpretability. A metric for evaluation was proposed using supervised data and the kurtosis measure, this metric achieved high correlation with manual scoring. Three topic models were compared in terms of the amount and quality of | Point graph and bar chart graphs displaying kurtosis of various bars. | Top 20 topics |

| | | | | | | | | | | |
|----|---|---|------|---|----------------------------------|---|---|--|--|---|
| | | | | | | 'musixmatch' and a dataset containing 8.598.630 social tag assignments, 'Last.fm'. | | unique themes. Finally, some applications of topic models for Music Information Retrieval are presented. | | |
| 25 | Trykowski, T. http://noisyvice.com/blog/rap-game-riff-raff-versace-lyrics | Rap Game Riff Raff Textual Analysis | 2013 | Undertook a textual analysis of Riff Raff's lyrical canon. We cleaned up every Riff lyric on Rap Genius: 116 songs and feature verses clocking in at over 20,000 words. We then threw the gigantic word document into a few fun experiments. | Rap/HH | Ragenius.com lyrics of Riff Rapps songs | 116 songs, 20000 words | Diagrammatic analysis of selected words - ice, candy, versace, ball, played for, rap game | Word cloud and word trees for selected words | Selected words - ice, candy, versace, ball, played for, rap game |
| 26 | (1)Wilkinson, A. http://www.newyorker.com/magazine/2013/04/01/rap-sheet-2 (2) Kyles, Y. http://allhiphop.com/2013/07/17/hip-hop-word-count-database-calculates-the-educational-level-of-popular-hip-hop-songs/ | (1) Rap sheet! 2) 'Hip Hop Word Count' Database Calculates The Educational Level Of Popular Hip Hop Songs | 2013 | Hip Hop Word Count allows Hemphill to calculate data like word count, average syllables per word, average letters per syllable, longest polysyllabic word, and Flesh and SMOG (Simplified Measure of Gobbledygook) scores to determine the educational level needed to grasp the meaning of a song's lyrics. It also takes into account artistic cleverness such as use of similes, metaphors, cultural references, and rhyme patterns when tallying its final number on a scale between 0 (illiterate) to 20 (post-graduate degree). | Rap/HH | Hip Hop Database including lyrics from 50K tracks and 3000 artists | 50,000 rap songs from 3,000 artists, from 1979 to the present | NA | NA | Educational level of lyrics |
| 27 | Zichermann, S. http://tspace.library.utoronto.ca/bitstream/1807/36081/1/Zichermann_Sandra_C_201306_EDD_thesis.pdf | The Effects of Hip-Hop and Rap on Young Women in Academia | 2013 | Investigates the rise of the cultures and music of hip-hop and rap in the West and its effects on its female listeners and fans, especially those in academia. Conducted a content analysis of 95 lyrics from the book, Hip-Hop & Rap: Complete Lyrics for 175 Songs by Spence. The songs had lyrics that repeated misogynist and sexist messages. Also conducted focus group on female students to see impact of lyrics. | Rap/HH | lyrics from the book, Hip-Hop & Rap: Complete Lyrics for 175 Songs by Spence. | Complete Lyrics for 175 Songs | Content analysis looking for 5 potential themes: Sexual exploitation and disrespect towards women, Ostentatious display of wealth, Glamorization of tobacco, alcohol, illegal substances, and weapons, Territory establishment, inclusion of derogatory and racially charged terminology. | None | Misogynist and sexist messages. |
| 28 | Davis, K. https://baylor-ir.tdl.org/baylor-ir/handle/2104/8206 | <i>Get Rich or Die Tryin': A Semiotic Approach to the Construct of Wealth in Rap Music</i> | 2011 | Signification and Representation of Wealth Themes - sex, living the good life, and respect - and signifiers (money, cars, attire, liquor and bling) as displayed in text of rap lyrics | Rap/HH | Top Billboard songs of last decade | 11 songs | "Hermeneutic" content analysis of songs to identify significations and representations with focus on major genres (e.g. Gangsta Rap) | None | Wealth - Materialism themes |
| 29 | Moddy, M. http://acjournal.org/journal/pubs/2011/spring/Moody-Ramirez.pdf | A Rhetorical Analysis of the Meaning of the "Independent Woman" In the Lyrics and Videos of Male and Female Rappers | 2011 | Combines feminist and critical cultural theories to explore the meanings of the "independent woman" in the lyrics and respective videos of male and female rappers. | Rap/HH | Analysis of the rap songs/videos yielded from search of songs in 2010 containing the keywords "women" and "independence" | Songs/videos of 6 rappers | A rhetorical analysis of collected songs, videos and video comments to compare and contrast perceptions of independence by male and female rappers and audiences. | None | Misogynistic language vs. images of independence with material wealth for women |
| 30 | Oware, M. https://www.researchgate.net/publication/225967661_Decent_Daddy_Imperfect_Daddy_Black_Male_Rap_Artists_Views_of_Fatherhood_and_the_Family | Decent Daddy, Imperfect Daddy: Black Male Rap Artists' Views of Fatherhood and the Family | 2011 | Exploratory research examines how black male rappers talk about motherhood, fatherhood, and parenthood. | Rap/HH | Sample includes platinum selling "rap" albums (selling at least one million copies) from 2004 to 2009. I employed the Recording Industry Association of America (RIAA) to help identify platinum selling rap albums. I distinguish "rap" albums from "hip hop" albums | 391 of rap songs from the years 2004 to 2009 | Contents analysis to understand how they talk about motherhood, fatherhood, and parenthood. | None | sexism, misogyny, violence, and hypermasculinity; attitude toward marriage and family; love, protection, and material support toward their mothers and children; relationships with their biological fathers; attitudes towards (biological) mothers of their children—baby mamas |
| 31 | Primack, B. et al. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3004676/ | Content Analysis of Tobacco, Alcohol, and Other Drugs in Popular Music | 2010 | Analyzed the 279 most popular songs of 2005 according to Billboard magazine. Two coders working independently used a standardized data collection instrument to code portrayals of substance use. Outcome Measures—Presence and explicit use of substances and motivations for, associations with, and consequences of substance use. | Variety including R&B/HH and Rap | Most popular songs from 2005 Billboard magazine | Lyrics from 279 songs | Content analysis of songs resulting in frequency analysis of references to substance abuse, motivations for use, as well as consequences. Detailed coding for these dependent variables are in the paper. The type of substance was noted along with coded descriptive information related to each song from Billboard's records, including song title, artist, album, song length in minutes and seconds, sex of singer, and primary song genre. Genre included 5 -- country, pop, R&B/hip-hop, rap, and rock). Chi square analysis used to determine differences between types, motivations and consequences. | None | Substance abuse |
| 32 | van Zaanen, M. and Kanters, P. | Automatic Mood | 2010 | Using a collection of lyrics and corresponding user-tagged | Variety | From 10K songs in Moody database, | Lyrics for 5631 songs | All docs of particular mood are combined together leaving 10 songs to compare with | None | Mood |

| | | | | | | | | | | |
|----|---|--|------|--|---------------|---|---|---|------|--|
| | http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.188.2073 See also: http://dc.ofai.at/browser?i=18914&y=all | Classification Using Tf*Idf Based on Lyrics | | moods, we build classifiers that classify lyrics of songs into moods. By comparing the performance of different mood frameworks (or dimensions), we examine to what extent the linguistic part of music reveals adequate information for assigning a mood category and which aspects of mood can be classified best. | | found songs with complete info and looked up associated lyrics. Used mood tags from Crayonroom (now defunct) | | respect to character count, word count, and line length. Computed td-idf and then ran TIMBL k-NN classifier to see what best predicted/distinguished mood up the 10 categories. | | |
| 33 | Binfield, M. https://repositories.lib.utexas.edu/bitstream/handle/2152/7555/binfieldm85962.pdf?sequence=2&isAllowed=y | Bigger Than Hip-Hop: Music and Politics in the Hip-Hop Generation | 2009 | Investigates the links between hip-hop music and culture and politics in the lives of audience members, exploring audience member's definitions of politics and community and examining the influence of hip-hop on these definitions. | Rap/HH | NA | NA | Ethnographic project that includes participant observation as well as in-depth interviews with self-identified hip-hop fans. Participant observation took place at two National Hip-Hop Political Convention conferences, in Austin at concerts, panel discussions, and other hip-hop oriented events, and online in an email listserv devoted to hip-hop and politics. | NA | Political messages |
| 34 | Lee, M. http://www.Overthinkingit.com/2009/08/11/textalyzing-born-to-run-straight-outta-compton-taylor-swift/ | <i>Textalyzing Albums: Word Frequency in the Lyrics of Born To Run, Straight Outta Compton, and Taylor Swift</i> | 2009 | Word Frequency in the Lyrics of 2 pop songs Born To Run, Straight Outta Compton, and Taylor Swift | Rap/HH | Top 10 words in 3 albums -Born To Run, Straight Outta Compton, and Taylor Swift | Top 10 words in 3 albums - Born To Run, Straight Outta Compton, and Taylor Swift | Free text analysis tool on UsingEnglish.com - eliminates stop words | None | Top 10 words |
| 35 | Oware, M. http://www.istor.org/stable/40282596?seq=1#page_scan_tab_contents | A "Man's Woman"? Contradictory Messages in the Songs of Female Rappers, 1992-2000 | 2009 | Employ content analysis of 44 songs from Billboard charts between 1992 and 2000. Find that female rap artists in sample include same themes as men - braggadocio, consumption of alcohol and drugs, and dissin' female and male competitors. | Rap/HH | Billboard Top 100 and Billboard Rhythm and Blues/Hip Hop Charts using songs on the charts 3+ months on chart. | Lyrics from 44 BB songs between 1992 and 2000. | Individual analysis and discussion for each of the five categories | None | Focus on 5 topics: braggadocio, consumption of alcohol and drugs, and dissin' female and male competitors. |
| 36 | Weitzer, R. and Kubrin, C. https://webfiles.uci.edu/ckubrin/Misogyny%20in%20Rap%20Music.pdf?uniq=fn1t7r | Misogyny in Rap Music: A Content Analysis of Prevalence and Meanings | 2009 | Assess the portrayal of women in a representative sample of 403 rap songs. Content analysis identified five gender-related themes in this body of music— themes that contain messages regarding "essential" male and female characteristics and that espouse a set of conduct norms for men and women. | Rap/HH | All rap albums from 1992 through 2000 that attained platinum status (selling at least 1 million copies). N= 130 albums. List from Recording Industry Association of America (RIAA). | The 130 albums contained a total of 1,922 songs. Using SPSS, a simple random sample of 403 songs was drawn and then analyzed. | Content analysis of every line coding whether misogynistic or not and then determine class of misogyny - (a) derogatory naming and shaming of women, (b) sexual objectification of women, (c) distrust of women, (d) legitimization of violence against women, and (e) celebration of prostitution and pimping. Frequency counts of songs by class. | None | Misogyny |
| 37 | Petrie, K., Pennebaker, J., and Sivertsen, B. http://homepage.syu.utexas.edu/homepage/faculty/pennebaker/reprints/beates.pdf | Things We Said Today: A Linguistic Analysis of the Beatles | 2008 | The goal of the current project was to apply two relatively new computerized text analysis methods to the lyrics of the Beatles. By analyzing the individual words in the collected songs written by the group members, we tracked the development of their music over time (1964, 1965-67, 1968-70) and compared the relative contributions of Lennon, McCartney, and Harrison. Overall, the Beatles' lyrics became darker, more psychologically distant, and less immediate over time. Paul McCartney's lyrical style proved to be more variable and broad ranging than either Lennon or Harrison. Using latent semantic analyses, Harrison's lyrics were more influenced by Lennon than by McCartney. Finally, the lyrics jointly written by Lennon and McCartney were mathematically more similar to Lennon's linguistic styles than McCartney's. | Beatles Songs | The lyrics of Beatles songs were downloaded from two Web sites (lyricsdownload.com/beatles-lyrics.html; beatles-lyricsarchive.com/). | 185 songs included 78 by Lennon, 67 by McCartney, 25 by Harrison, and 15 by the Lennon-McCartney collaborations. | Two general strategies are available to analyze word usage—a word count method and a word pattern analysis. Used to answer the following questions: 1-To what degree was there a shift in emotional tone and cognitive analysis over the lifetime of the Beatles—for example, to what degree did the band become more negative and complex? 2-How did the Lennon-McCartney songs (which were jointly written) differ from those written solely by Lennon and McCartney as well as George Harrison? 3-To what degree are the differences among the writers' lyrics a function of their content or their linguistic styles? Relied on the text analysis program, Linguistic Inquiry and Word Count (dividing into psychological content dimensions or word categories) for second analysis. The second text analysis approach is a word pattern analysis which has emerged from the artificial intelligence community. Word pattern strategies mathematically detect "bottom-up" how words co-vary across samples of text. Latent Semantic Analysis used to determine within artist and between artist covariation. | None | Emotional and psychological content of songs |
| 38 | Mahedero, J., Martinez, A., and Cano, P. https://www.researchgate.net/publication/221573745_Natural_language_processing_of_lyrics | Natural Language Processing of Lyrics | 2005 | Explores the use of natural language processing of song lyrics for purposes of language recognition, structure extraction, thematic categorization, and similarity searches. | Variety | Song lyrics obtained by crawling different lyric sites (e.g. lyrics.com or lyrics4u.com) | 500 songs for language recognition; 30 songs for structure extraction, 125 songs for thematic extraction, and similarity comparisons based on different versions of a small set of songs. | Varying analysis depend on the specific use of NLP: (1) language recognition employed Ted Dunning's statistical identification using bi/trigrams; (2) structure extraction used a specialized algorithm designed to automatically identify the intro, chorus, verses, bridges and outro; (3) thematic categorization employed Naïve Bayes; and (4)similarity searches used cosine distances among "idf" word weights for songs. | None | Song structure and themes (love, violence, protest, Christian and drugs). |
| 39 | Logan, B., Kositsky, A., and Moreno, P. | Semantic Analysis of Song Lyrics | 2004 | Explores the of song lyrics for automated indexing of music. Utilizes standard text processing | Variety | Lyrics primarily from azlyrics.com for the 400 artists | Lyrics for 15,589 songs associated with | Employed Probabilistic Latent Semantic Analysis (PLSA) to cluster lyrics based on the words that appear most frequently. The assignment of | None | Word Frequencies |

| | | | | | | | | | | |
|----|--|--|------|---|------|--|--|---|---|----------------|
| | www.researchgate.net/publication/4124757_Semantic_analysis_of_song_lyrics | | | to characterize semantic content and discover genre clusters among artists. | | in the uspop2002 dataset | the artists in uspop2002 dataset as well as from a survey (with 10 +K comparisons) designed to determine the similarity among artists. | artists to clusters are compared to genre assignment take from the All Music Guide (allmusic.com - note that this site still exists) and a survey (with 10+K comparisons) designed to determine the similarity among artists. | | |
| 40 | Peng, R. and Hengartner, N. http://www.biostat.jhsph.edu/~rpeng/papers/archive/authorship-tas2-final.pdf | Quantitative Analysis of Literary Styles | 2002 | Overview and brief history of the analysis of literary styles. Use canonical discriminant analysis and principal component analysis to identify structure in the data and distinguish authorship. Assume: (1) the style of an author remains the same throughout his/her life;(2) successive occurrences of function words are independent. Neither assumption tends to hold in practice. | None | Raw data for this study were obtained from Internet websites such as Project Gutenberg. Multiple works for each author were downloaded in text format and processed. The titles and website URL are listed in Appendix A | 9 authors and 59 works | Divide each author's work into 1700 words. For each block, we tabulated the frequency of the 69 words chosen from the Miller-Newman-Friedman list of function words used. Canonical discriminant analysis was used to provide dimension reduction and graphical displays of the differences between authors (canonical vector plots). Also, CDA was useful for identifying key function words which were most effective at discriminating between authors. The key words were identified by examining plots of the loadings for each function word. Function words were chosen as the unit analysis because they are highly variable between authors, abundant, and easy to count and identify. Function words include: (a been had its one the were all but has may only their what also by have more or then when an can her must our there which and do his my should things who any down if no so this will are even in not some to with as every into now such up would at for is of than upon your be from it on that was) | Scatterplots of Components within Authors and between authors | Function words |