Using Factor Analysis and Quantitative Content Analysis to Detect Themes in Media Texts: A Comparison of Pre- and Post-9/11 Song Lyrics

Brenda L. Osuna[1] PhD, Reagan Rose[2]
[1]Information Technology Services, University of Southern California,
[2]Dornsife College, University of Southern California

Abstract
Media, social networking, and communications data is of vital importance to every sector of business, industry, government, and academia. Research studying the effects of media in these industries impacts public policy and shapes decision-making in personal and business spheres. Creative application of statistical methods can potentially reduce bias and increase precision in measurement of media content. Osuna (2011) analyzed a sample of music lyrics (*N*=300) using Linguistic Inquiry Word Count and principle components analysis. Factor scores for each lyric were obtained on the resulting dimensions. Each lyric was measured on a continuous scale for thematic content. The validity of the model was tested using Latent Semantic Analysis. Here, the utility of this method is demonstrated through an analysis of samples of pre- and post-9/11 music lyrics to obtain factor scores. Then, using a t-test for each factor dimension, the lyrics are examined for differences. Results indicate that the method is worthy of more exploration and validity testing. We recommend further study of methodology and measurement techniques using statistical methods applied to media content.

Key Words: factor analysis, content analysis, song lyrics, text analysis, psychology of music

1. Introduction

Data mining of themes in media texts has become of increasing interest. The present research represents the first application of a method developed specifically for this purpose (Osuna, 2011). Using principal components analysis (PCA) to reduce the number of psychological dimensions in a previously validated content analysis method (LIWC, Pennebaker, Booth, & Francis) Osuna designed a method to detect and quantify psychological themes in popular songs that can easily be applied to other types of media texts.

For this study, it was hypothesized that popular song lyrics from before and after the events of September, 11 2011 would show significant differences on many psychological dimensions; hence, we chose a sample of these songs to provide early evidence for the utility of the Osuna method.

2. Method

A random sample (*N*=200) of pre- and post-9/11 songs was obtained from a dataset containing all the songs from the *Billboard Hot 100* for the years 1990-2010. To account for estimated production time lag, the songs were sampled from two 4-year periods: August 1997 - August 2001 for the pre-9/11 sample and from August 2006 - August 2010 for the post-9/11 sample. Songs with non-English lyrics were excluded from the sample resulting in a final sample of *N*=197.

Figure 1: Word clouds depicting the 100 highest frequency lyrics for both the Pre- and Post-9/11 samples

Song lyrics were collected from the online lyrics databases lyrics.com and metrolyrics.com. Lyrics were then analyzed using quantitative content analysis software that detects and quantifies the type and amount of 72 psychological dimensions present in text (Linguistic Inquiry Word Count [LIWC], 2007, Pennebaker, Booth, & Francis). Component coefficients were obtained from Osuna's (2011) PCA that reduced the number of psychological factors in LIWC to 6 factors: Anger and Impulsivity; Negative Emotions and Melancholy; Death, Religion, and Isolation; Daily Concerns; Positive Feelings and Romantic Relationships; and Inner Thoughts, Reflections, and Desires. Lyrics were then factor scored by multiplying the component coefficients for each factor by the corresponding LIWC word count for each song. Data were explored and extreme outliers were removed from the dataset. An independent groups t-test was conducted between pre- and post-9/11 samples for each factor.

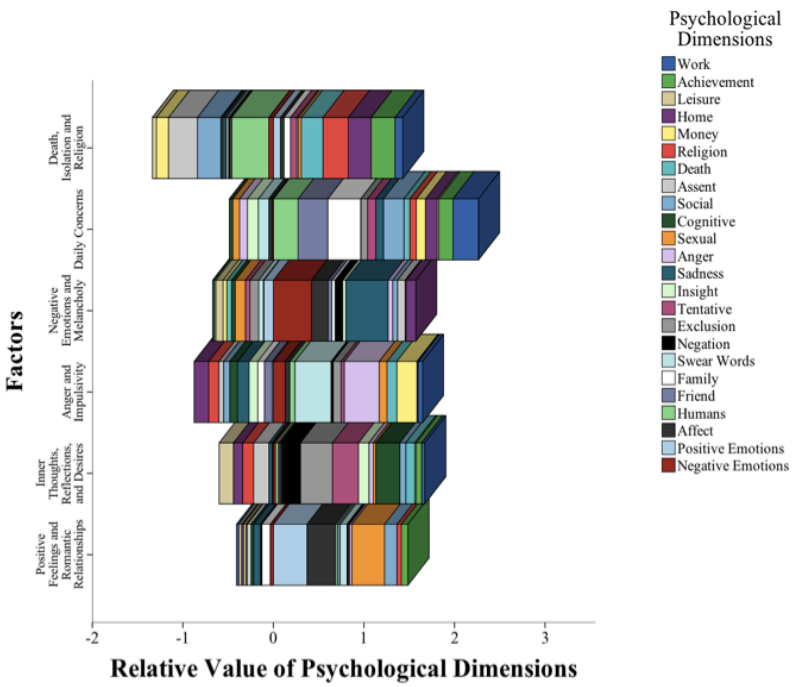


Figure 2: Relative value of each psychological dimension in each factor

3. Results

There was a statistically significant difference between pre- and post-9/11 lyrics on 4 of the factors: Anger and Impulsivity ($t(193)=-6.25$, $p < .01$), Negative Emotions and Melancholy ($t(193)=-3.66$, $p < .01$), Daily Concerns ($t(195)=-3.11$, $p < .01$), and Death, Religion, and Isolation ($t(194)=-3.59$, $p < .01$). The direction of significance for all factors that showed a difference between pre- and post-9/11 samples indicates that popular song lyrics reflect an increase in content related to Anger and Impulsivity; Negative Emotions and Melancholy; Daily Concerns; and Death, Religion and Isolation.

There was no significant difference on either the factor representing Positive Feelings and Romantic Relationships or the factor representing Inner Thoughts, Reflections and Desires. This indicates that that the lyrical content of popular songs did not change after the events of 9/11 in these areas.

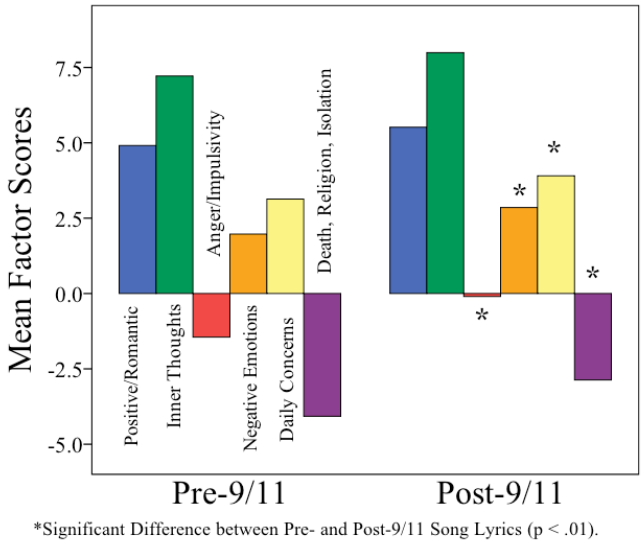
4. Conclusions

Our study affirms that significant differences between pre- and post-9/11 song lyrics do exist and these differences are reflective of the sobering realities of the post-9/11 era. Post-9/11 popular songs reflect an increase in lyrical content related to negative psychological dimensions, increased concerns, and religion. Conversely, the importance of romantic relationships, positive feelings, and inner reflections in post-9/11 songs does not differ from songs of the pre-9/11 era. The method used to detect these differences shows early evidence of utility and validity. We plan further studies to statistically evaluate the convergent and discriminant validity of this method. Because this method provides unbiased quantitative assessment of content in text, there is potential for its use

in applied research in a variety of fields that currently use less precise measures of psychological content analysis.

Table 1: Means and standard deviations for factor scores

	Pre-9/11 Song Lyrics (*N*=99)	Post-9/11 Song Lyrics (*N*=98)
	Mean (SD)	*Mean (SD)*
Factor 1: Positive Feelings and Romantic Relationships	4.70 (2.36)	5.29 (2.82)
Factor 2: Inner Thoughts, Reflections, and Desires	7.39 (2.70)	7.99 (2.32)
Factor 3: Anger and Impulsivity	-1.45 (1.36)	-.23 (1.35)
Factor 4: Negative Emotions/Melancholy	1.90 (1.83)	2.79 (1.55)
Factor 5: Daily Concerns	3.13 (1.66)	3.90 (1.81)
Factor 6: Death, Religion, and Isolation	-3.91 (1.91)	-2.87 (2.15)



*Significant Difference between Pre- and Post-9/11 Song Lyrics ($p < .01$).

Figure 3: Mean factor scores for each factor

References

Osuna, B. L. (2011), "Psychological Effects of Music with Lyrics: A Methodological Study", *Proquest Dissertation and Theses Database,* UMI No. 54580.

Billboard (2013), "The Hot 100", Retrieved from http://www.billboard.com/charts/hot-100

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007), LIWC2007: Linguistic inquiry and word count (Version 1.0.8). Austin, Texas: liwc.net.