

From *The Beatles* to *Bad Bunny*: Sexism in popular music through an automated text analysis

Keywords: music, sexism, lyrics, machine learning, natural language processing.

Extended Abstract

Motivation and Goals

Sexism against women, as a discrimination practice, usually implies violence, harmful acts, or reinforcement of stereotypes, and persists as a deeply-rooted problem. Cultural production, particularly music, can be seen as both a reflection and a mirror of the society in which it is embedded. The prevalence of sexism in music serves as a reflection of the level of sexism that exists within our society, but can also influence it. In particular, song lyrics could have the ability to impact human behavior in a profound manner. Hence, it is a matter of concern that messages in lyrics might be sexist.

We examine the progression of sexism against women in music lyrics over the past six decades (1960–1922) in Spain, looking at the most listened to songs. While sexism in music lyrics have been widely documented through manual analysis (e.g., [3]), and automated content analysis has been used to study gender bias in lyrics (e.g., [2]), we undertake an automated, large-scale analysis of sexism in lyrics specifically. As an initial step towards an intersectional analysis, we also problematize the power interrelation between sex and race in lyrics, detecting whether there are any references to individuals who have been racialized among the lyrics. When dealing with large amounts of songs, manual analytical methods are often unfeasible. Therefore, to perform a large-scale analysis, we used automatic text classification to categorize the lyrics of about five hundred songs into sexist/non-sexist and racialized/non-racialized categories, as well as identifying broad types of sexist speech present in the lyrics.

Methods

To create the dataset, we collected the titles, artist names, and language of the most listened to songs in Spain on each year from 1960 to 2022, using local sources such as *Los 40 Principales* and *Los Superventas* for the initial decades, and *Spotify* for the most recent decades. Then, we used the *Genius API* (<https://genius.com/developers>) to download the lyrics of all the collected songs. Lyrics were then segmented into paragraphs for manual annotation, which subsequently served as the training data for our final classifier.

To train the model, a sample of lyrics paragraphs (i.e., verses and choruses) were labeled as sexist/non-sexist and racialized/non-racialized. To make labeling more efficient, each version of the model was used to stratify the sample of data to label for the next version iteratively, for three model iterations. To represent lyrics text as vectors for modeling, we used Language-Agnostic Sentence Representations (LASER) [1], which are multilingual, combined with a Long Short-Term Memory (LSTM) based model.

In parallel, we carried out two crowdsourcing tasks to gather information on how sexism is represented in lyrics. For this task, we focused solely on songs in Spanish, as they were the most prevalent in sexist songs. The task was to classify the verses in the paragraphs as sexist

or not based on a given definition, and to identify various categories of sexism including role stereotyping, women objectification, “slut shaming,” among others.

Results

We first examine the evolution of sexism over the course of each decade. Figure 1 a) illustrates the percentage of songs having lyrics with sexist expressions. We can see that sexism has progressively worsened in recent decades, except for a peak in the 1970s. Throughout the decades leading up to the 2000s, sexism remained relatively stable. However, a critical finding of this analysis is the sharp increase in sexism during the past two decades. Our results indicate that roughly 82% of popular songs contain sexist expressions.

However, as the final decade analyzed includes only the years 2020, 2021 and 2022, we opted to further examine the progression of sexism in smaller subsets of years for greater precision. Figure 1 b) depicts the evolution of sexism over 5-year periods since 1960. Similarly to the previous findings, the prevalence of sexism in lyrics remained consistently – between 37% and 50% of the songs – until 2000. Nevertheless, in contrast to the previous analysis, the trend persisted between 2000 and 2010, with 37% to 53% of lyrics being identified as sexist, and it increases during the last ten years to 82% as previously stated.

Additionally, the study revealed that an overwhelming proportion of songs that contain racialized references also exhibit sexist undertones. This suggests a possible association between racialization and perpetuation of sexist attitudes in lyrical content. Qualitatively, we note that among the most frequent words in sexist verses and choruses we find those related to pleasure, women, women’s bodies, and sensual movements. In general, these phrases are mainly concerned with sexual acts or circumstances, and use imperative verbs, ordering women to perform certain acts.

Overall, the findings indicate that sexism has always existed in the lyrics of the most popular songs in Spain. However, sexism in lyrics today is more widespread than in previous decades. A conjecture we want to study is whether this happens due to changes in the theme/topic of songs (e.g., a fraction of songs related to romantic love or sex has increased). This study may lead towards impacting applications that have the potential to mitigate sexist biases that are prevalent in society. Moreover, it can potentially be used to change how broadcasting apps make recommendations, considering the sexist and/or racialized bias in the songs that are promoted by algorithms in certain platforms.

References

- [1] Mikel Artetxe and Holger Schwenk. “Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond”. In: *Transact. of the Assoc. for Comp. Linguistics* 7 (2019), pp. 597–610.
- [2] Ashish Gupta et al. “Brave Men and Emotional Women: Analyzing Gender Bias in Bollywood Songs”. In: *Proc. of Int. Conf. on Comp. Int. and Data Eng.* Springer. 2021, pp. 143–153.
- [3] Andrew P Smiler, Jennifer W Shewmaker, and Brittany Hearon. “From “I want to hold your hand” to “promiscuous”: Sexual stereotypes in popular music lyrics, 1960–2008”. In: *Sexuality & Culture* 21.4 (2017), pp. 1083–1105.

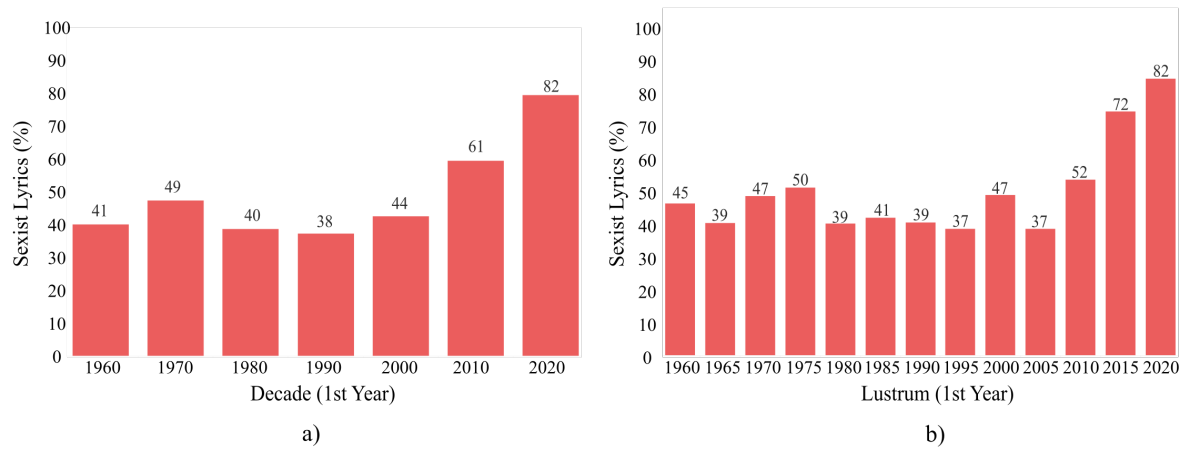


Figure 1: a) Percentage of sexist songs by decade. b) Percentage of sexist songs by lustrum.