Chapter 47
Computational Analysis of Latin Music Songs Through Tokenization. Case of Female Artists and Reggaeton

Angel Torres-Toukoumidis, Cristian Orlando Coronel Quezada, Jenny Pontón, and Isidro Marín-Gutiérrez

Abstract The following research focuses on the analysis of the computational analysis of songs from female reggaeton artists. For this purpose, 641 songs from 12 singers were reviewed, the information obtained was extracted using a Python script with Selenium library, thus allowing access to the compilation of the lyrics, then to establish patterns and similarities, the R software was applied, complemented with the Voyant Tools tool. The results show that the songs interpreted by women contain an abundant lexicon, which in turn is augmented with anglicisms that make their legibility difficult. At the same time, the positioning of women in a dominant, decisive and outstanding role with respect to the actions involved in the typicality of urban-Caribbean music is demonstrated. In conclusion, women as singers in reggaeton have managed to take over a space homogenized by patriarchy, thus fostering a diversification of the musical entertainment industry with other profiles, identities and lyrics, thus building a diversification of this subculture.

47.1 Introduction

Reggaeton, a musical genre from the Caribbean context, has become a relevant component of urban culture [1]. Its great popularity and impact on the entertainment industry are evidenced by the results of its millions of downloads, streaming revenues and its positioning in the top positions of Billboard capturing with special attention

A. Torres-Toukoumidis (✉)
Universidad Politécnica Salesiana, 010105 Cuenca, Ecuador
e-mail: atorrest@ups.edu.ec

C. O. C. Quezada
Universidad Técnica Particular de Loja, 110107 Loja, Ecuador

J. Pontón
Facultad Latinoamericana de Ciencias Sociales, FLACSO, 170201 Quito, Ecuador

I. Marín-Gutiérrez
Universidad de Sevilla, 41004 Seville, Spain

the uniqueness of the processes of the music market of this genre by the academic community from different multidisciplinary notions [2]. In fact, as a symbolic case is "Despaci-to" of 2017, a song performed by Luis Fonsi and Daddy Yankee that managed to be the music video to surpass the record of 7 billion views on streaming platforms [3] remarking a milestone for the globalization of these rhythms through the digitization and decentralization of their consumption.

While it is true that there is a multiplicity of research on the phenomenon of reggaeton, it should be noted that this research aims to focus on the field of communication, hence several issues should be specified; reggaeton lyrics transmit sexual messages and configure the role of women as passive agents of men's actions [4] complementing this idea, the discourse of reggaeton music videos was also analyzed, maintaining the same premises of female stereotyping as sexual objects, dehumanized and superficial with zoomorphic characteristics [5, 6], but beyond the evident preponderant misogyny, there are also cases in which the uninhibited, liberal and modern woman is represented [7]. Continuing in this line, the feminist movement has also appropriated the genre with lyrics and interpretations defending the individual freedom of women and the vindication of their enjoyment serving as a tool for the empowerment of their rights [8].

In contrast to the sexist message established in the lyrics and videos, the verbal framing of the media discourse on this musical genre in Medellin has incurred in the construction of an idealized geographical space for the Colombian music industry as an incentive for the prevention of vandalism and drug use [9]. Regarding this issue, reggaeton has also intensified a Caribbean identity at a transnational level, so much so that it has overcome the barriers circumscribed by Cuban state agencies allowing the dissemination of musical videos, progressively breaking through censorship by being assumed as a cultural expression despite being incorporated in the well-known "digital packages" coming from abroad [10].

Within the communicative context, another important issue taken into account by the academic literature is related to social networks. In addition to the marginalization and denigration of women reiterated with a hegemonic discourse via social networks [11], online platforms have provided a space for the promotion of fan communities linked to this musical genre, inviting participation, recognition, preservation and praise of reggaeton [12]. To culminate, reggaetoneros are media profiles with a high level of power and influence on social networks, particularly, Twitter, Facebook, Instagram with the echo of Spotify transcending with their image and language in the sympathizers of this rhythm, being models for their listeners [13].

In summary, the complexity of the communicative ecosystem exposed from reggaeton constitutes an incipient area of academic exploration. Therefore, the innovative character of this research is to place women at the core of this analysis, answering the following questions: What are the songs performed by female artists about? What role does the female gender play in the discourse of the lyrics performed by reggaeton artists? Are reggaeton artists sexualized in their songs or, on the contrary, do they sexualize men?

47.2 Methodology

The general objective of this research is analyzing the linguistic content of the songs of female reggaeton artists. Specifically, the following specific objectives have been established:

(a) Categorizing at the semantic level the songs performed by reggaetoneras.
(b) Identifying female representation in the discourse of songs performed by reggaetoneras.

In order to achieve the stated objectives, a qualitative approach based on content analysis was applied, which is defined as an approach that allows understanding reality through categorization of metatextual properties and characteristics [14]. This methodological design derives the singularity of being articulated under a linguistic demarcation based on the syntactic-denotative level [15]. In other words, the aim is to analyze the meaning of the words used in the texts obtained by recognizing the multiple variants and lexical combinations adjusted to the univocal reality of the speaker.

For this reason, this exploratory interpretative evaluation has established documentation as the main data collection tool. For this purpose, 641 songs were selected from 12 of the most recognized Ibero-American reggaeton artists according to their positioning in the top lists of Spanish radio programming -Los 40 principles as the most important selection criteria. The extraction has been done through Python with the Selenium library, with this basically the script is designed to enter Google, put in the search engine the artist + song + lyrics. For example: Karol G + Tusa + letter, once you searched, the lyrics of the song were presented, and the script automated the copy of the lyrics and saved it in a text file.txt. The following ordered Table 47.1 gives the artist and the number of songs belonging to the song.

Corresponding to the 641 songs obtained, the analysis was applied to 240,686 words from which the review was organized in 3 procedural phases according to the specific objectives presented. The first phase focused on a semantic-denotative review by means of the Voyant-tools tool in which the reiteration of terms used in the songs is glimpsed, knowing their tenor and frequency. In the second phase, the aim is to determine the feminine presence in the discourse, that is, to explain the self-representation in the songs through the codification of feminine stereotypes exposed by Almansa-Martínez and Travesedo-Rojas [15], and finally, the third phase, we also proceed to review the lyrics of the songs, but this time assessing the sexualization, evaluating whether there is a certain alignment with heteronormative male fantasies [16] by making erotic and provocative allusions in their lyrics [17].

Under this research, the second and third phases, considering the main objective of the research where the focus lies in knowing the content of the urban reggaeton genre from the perspective of the woman as an interpreter, we proceed to conduct a thorough analysis of the relationship through text mining.

The computational analysis of text is called "tokenizing" (tokenize), which corresponds to the activity of extracting units of analysis from a body of writing, which in

Table 47.1 Reggaeton artists
and number of songs

Artist	# of songs
1. Karol G	94
2. Becky G	84
3. Danna Paola	88
4. Ivy Queen	79
5. Leslie Grace	51
6. Natti Natasha	48
7. Nicki Nicole	44
8. Mari Becerra	38
9. Lola Indigo	37
10. Farina	32
11. K-narias	31
12. La Sista	15
Total	641

our case suggests the consolidation of songs by female performers. The individual words that compose it ("tokens") are extracted in order to quantify in some way its attributes or characteristic elements of this genre.

In our dataset, all songs are coupled into a single corpus, where sentences are separated by paragraphs per artist and show a running trend of events. We can use different perspectives of analysis that will be described in the study process. At the end, the word cloud generated by the behavioral similarities is contrasted and adjusted to a key theoretical element within this study. The following procedure is carried out using R SOFTWARE 4.1.3., by applying sentence reduction and homogeneity clustering algorithms.

47.3 Results

Evaluating the 641 reggaeton songs with 41,767 paragraphs and 240,686 pa-words, the length of the most representative words has an average of 6 letters in which a narrative is prescribed where personal pronouns predominate 9346 times with "yo, tu y el", anglicisms 2095 times with "baby, know, yeah, Flow, wanna" (Fig. 47.1).

As for the length of the most representative words, the average is 6 letters. This shows that the aim is to include short words in the lyrics of songs that facilitate the expression of the message exposed by the totality of the song, while the totality of the song lyrics contains an average of 795 words (Fig. 47.2).

On the other hand, per sentence, there is an average of 119.6 words with a lexical density of 0.060, i.e. this measure shows the relationship between the length of words and the number of distinct words. Reggaeton songs have a lexical density of 6%, i.e., 6 out of 100 words are repeated.

Fig. 47.1 Most repeated words in reggaeton songs


Fig. 47.2 Number of words in reggaeton lyrics
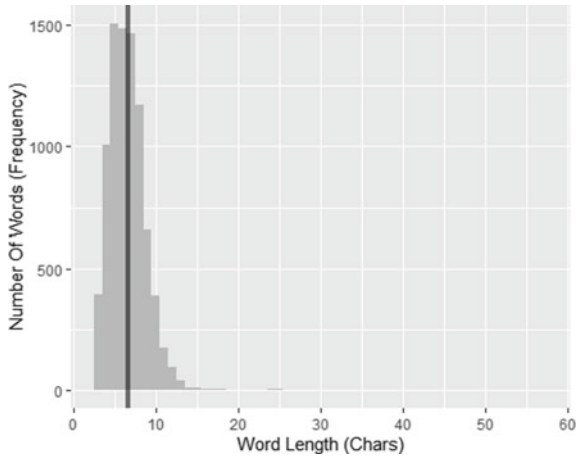
In correspondence to the readability index, the lyrics contain a 5.44, in other words, this means that it is below average and exhibits a certain degree of difficulty in reading.

The total number of words for the analysis of behavioral homogeneities is 237,438 in which a first filtering process was carried out, which revealed that there are a number of words reiterated more frequently than others (Fig. 47.3):

These data show that the lexicon of the songs performed by reggaetoneras is extensive, only 2 words are repeated more than 1000 times, while at the other extreme, 4992 words were used only 5 times. This clearly shows that in contrast to the beliefs held within the collective imagination that reggaetón is a genre with a limited vocabulary, the results show that reggaetón songs contain varied discourse options in their musical expression.

Fig. 47.3 Categories and
word frequency
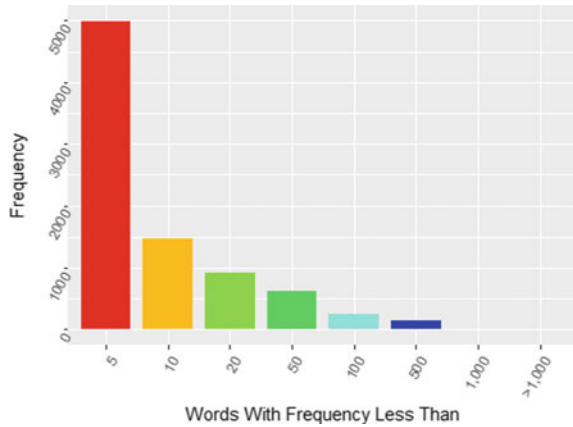
For the second purification, prepositions and basic words were omitted, generating a dataset to be able to investigate behaviors and emotions expressed in the songs. The table gives a top of the most repeated words in the reggaeton genre interpreted by women (Table 47.2).

The words with more than 1000 repetitions are <te> and <quiero> , on the opposite side, there is the appearance of single terms denoting in a particular way that most of them use the z. Other patterns obtained are again the combination with anglicisms, expression of actions and emphasis to the present, adding words like <corazon> with 507, <contigo> with 365, <tengo> 567, <gusta> 387, <estoy> with 491, <hoy> with 432, <ahora> with 493, <voy> with 521, or <estoy> 394.

When generating a second debugging, extracting some sentences from the sample distinguishes the following: Becky G, song 24/7.

a. "Te amo solo y de fondo suena un ritmo violento"

Table 47.2 Words in
reggaeton songs

Words	Freq
Te	1932
Quiero	1226
Soy	953
Amor	909
Baby	810
Mí	713
Zeta	1
Since	1
Zoom	1
Zooting	1
Zuchinni	1
Zumba	1

b. "Mi cuerpo se va moviendo y te pone' content"
c. "Hace unas horas, no te conocía"
d. "Era de noche y ahora de día"
e. "Yo que decía que no bebería"
f. "Y sigo aquí contigo todavía"

The selected phrases clearly demonstrate that there is a strengthening of the feminine image in direct and personal decision-making. Contrary to what was demonstrated in the theoretical framework that exposes the songs of reggaetón mascu-lino artists, the self-designation directed toward zoomorphic characteristics is absent in most of them, except for phrases such as: "Quítate tú que llego la caballota; La Perra, La Diva, La Potra; La mami que tiene el tumbao" by Ivy Queen feat by Don Omar. While it seeks to highlight her attitude and physical attributes from a dominating role. From this perspective, a mainly denotative language is wielded toward dance, enjoyment, entertainment and courtship, keeping the woman's decision as the last word.

Yo quiero bailar, tú quiere' sudar

Y pegarte a mí, el cuerpo rozar

Yo te digo: "sí, tú me puedes provocar"

Eso no quiere decir que pa' la cama voy

Ivy Queen (2003)

47.4   Conclusions

Regarding the general objective, analyzing the linguistic content of the songs of reggaeton artists. A total of 641 reggaeton songs performed by women were reviewed, demonstrating an important differentiation from those presented by male artists, particularly, the elaboration of this semantic descriptive review has provided significant contributions, among them, averages of most used words, length of paragraphs and readability index are calculated, showing that the genre contains an extensive lexicon generating a creative multiplicity of combinations of ideas. Secondly, the stereotypical notion of women is distanced through a predominant discourse where the feminine gender is prescribed with an active and dominant role, modifying the power dynamics in the flirtation.

Regarding the categorizations and discursive representation, two causal paths can be glimpsed: the first, assuming that the stigmatized perspective of self-perception changes in the lyrics of the songs, giving license to transfer the eroticism of this youth subculture toward a vision-oriented depatriarchalization, providing a spur to women's autonomy [18, 19]. On the other hand, the second causal path is summarized in the diversification of the hegemonic culture with the purpose of obtaining an economic profit, similar to what happened with the change from the physical format to the digital format [20], but in this case, optimizing the public and the interested parties of this genre toward the female voice. It is recommended that future researchers delve deeper into this dichotomy.

This analysis also admits another reflection, the first one referring to false empowerments [21, 22], that is, "reggaetoneras" still conceive some stereotypes accepted by their appearance and patterns coming from established social structures. Are lyrics still written by men? Are lyrics still produced by men? Although this implies only an initiative labeled as affirmative action measures, the fact of being included in the music industry and obtaining a space to present themselves as artists can be seen as a first step to begin to generate some influence and gradually modify the common concurrences set by the collective imaginary toward women in this musical genre.

Similarly, it is recommended that future research should analyze from a critical and qualitative point of view, the repercussion of the lyrics and their transmediality formats, i.e. that music videos, concerts, new audiences, among others.

References

1. Morales, C., Cortés, P.: Empoderamiento femenino y cultura urbana en las series Bravas y La Reina del Flow. Miguel Hernández Commun. J. **12**, 587–609 (2021)
2. Salvado, M.A.: Rasgos estilísticos del reggaetón mainstream, una aproximación desde la producción musical. Etno: Cuadernos de Etnomusicología **15**(2), 130–156 (2020)
3. Rivera-Rideau, P., Torres-Leschnik, J.: The colors and flavors of my Puerto Rico: mapping "despacito"'s crossovers. J. Pop. Music. Stud. **31**(1), 87–108 (2019)
4. García, M.U.: El reggaetón, invitación al sexo. Análisis lingüístico. Temas de comunicación (20), 141–160 (2010)
5. Murillo, C.B.: Los estereotipos femeninos en los videos musicales del género reggaetón: una cuestión de género. Estudios (20), 184–198 (2007)
6. Morales, C.D.R.M.: Una propuesta para el análisis de los estereotipos femeninos en los videoclips de reggaeton: Caso práctico de los cuatro vídeos más vistos en 2018 en YouTube. Vis. Rev. Int. Vis. Cult. Rev. **7**(1), 13–26 (2020)
7. Gallucci, M.J.: Análisis de la imagen de la mujer en el discurso del reggaetón. Opción **24**(55), 84–100 (2008)
8. Araüna, N., Tortajada, I., Figueras-Maz, M.: Feminist reggaeton in Spain: young women subverting machismo through 'perreo.' Young **28**(1), 32–49 (2020)
9. Mykhalonok, M.: La Capital mundial del reggaetón: El framing verbal de Medellín en el discurso mediático en línea. Ricercare **13**, 56–71 (2020)
10. Pereira, S.L., Soares, T.: Reggaeton in Cuba: censorship, ostentation and cracks in media policies. Palabra Clave **22**(1), 98–112 (2019)
11. Ellis, V.D.: "¿Dónde están las yales?" Reggaetón and womanhood in the welfare Island. J. Gend. Sex. Stud./Revista de Estudios de Género y Sexualidades **46**(1–2), 195–214 (2020)
12. Chalk, L.: Representing reggaeton through fans' online community archives. Transformative Works Cult. **37** (2022)
13. Moreno López, G.: A multimodal discourse analysis of the most viewed reggaeton video on youtube by the LIV super bowl halftime show. Open J. Stud. Arts **3**(2) (2020)
14. Campos, M.M., Mujica, L.A.: El análisis de contenido: una forma de abordaje metodológico. Laurus **14**(27), 129–144 (2008)
15. Almansa-Martínez, A., de Travesedo-Rojas, R.G.: El estereotipo de mujer en las revistas femeninas españolas de alta gama durante la crisis. Rev. Lat. Comun. Soc. **72**, 608–628 (2017)
16. González-Carrión, E., Torres-Toukoumidis, A., Chaljub-Hasbún, J., Rivera-Rogel, D.: Género y estereotipación de las videojugadoras. Caso google imágenes. In: Grana, R. (eds.) Discursos, mujeres y artes. ¿construyendo o derribando fronteras? pp. 2242–2260, Dykinson (2021)

17. Merlyn, M.F.: Dime lo que escuchas y te diré quién eres. Representaciones de la mujer en las 100 canciones de reggaetón más populares en 2018. Feminismo/s (35), 291–320 (2020)
18. Alcazar, M.V.: Feminismo, juventud y reggaeton: cuando las mujeres cantan y perrean. Vitam. Revista de investigación en Humanidades (3), 36–57 (2018)
19. Dediós, P.J.V.: Reguetón de mujeres para mujeres: dos caminos de empoderamiento. Antec: Revista Peruana de Investigación Musical **5**(1), 91–109 (2021)
20. Rivera, M.M.: Hate it or love it: global crossover of reggaetón music in the digital age. University of Illinois at Urbana-Champaign (2014)
21. Sáez, G., Valor-Segura, I., Expósito, F.: ¿Empoderamiento o subyugación de la mujer? Experiencias de cosificación sexual interpersonal. Psychosoc. Interv. **21**(1), 41–51 (2012)
22. Lamb, S., Peterson, Z.D.: Adolescent girls' sexual empowerment: two feminists explore the concept. Sex Roles **66**(11), 703–712 (2012)