Riri in the Wild Zone

A Distant Reading of Sapphic Urban Latin Music

Laura Isabel Sastoque Pabón

Table of contents

1 Introduction

In recent years, distant reading methods have earned popularity for the analysis of literary texts. These methods hinge on the transformation of text into features— data that can be quantified and visualized using computational tools.[1] Often framed in methodological opposition to close reading, distant reading has been scrutinized for its impulse to automate interpretation. However, scholars such as Katherine Bode have theorized that distant reading does not stand in opposition to close reading, but is rather a conceptual legacy of the latter, which prioritizes the text as the ultimate source of meaning.[2] Furthermore, scholars have called attention to distant reading methods for their potential to expand existing literary canons through a

---

[1] Maciej Eder, et al. "Stylometry with R: A Package for Computational Text Analysis," *The R Journal 8* (January 1, 2016): 107–21.

[2] Katherine Bode. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly* (Seattle) 78, no. 1 (2017): 77–106.

wider, more transparent inquiry of literary material.[3] So far, the majority of applications of distant reading in the digital humanities have had literary texts as their objects of inquiry—novels, poems, even television scripts.[4] However, what happens when the objects of a distant reading study are not commonly considered "literary"? What happens when these texts are, in fact, not even meant to be read, but instead heard, or danced to?

Few studies have ventured into the realm of music wearing a distant reading toolbelt. Those that do, contain little critical reflection on the revelations that this textual manifestation brings to the method.[5] This essay aims to conduct a distant reading of sapphic urban Latin music with an open, yet critical lens that will provide a deeper view into the pertinence of this method. In recent years, urban Latin music—known for its highly sexualized lyrics and male-dominated industry—has transformed.[6] Through hit songs such as Young Miko's "Riri," which speak about woman-for-woman desire, a subset of sapphic artists is earning notoriety.[7] In a genre where the hyper-sexualization of women is a convention, it is pertinent to interrogate whether sapphic songs— particularly, at a lyrical level— frame desire in a similar way to straight male artists or create a new lexical tradition that breaks pre-existing conventions.

By conducting analyses such as text mining and stylometry, this study aims to test distant reading's ability to understand the appropriations and transformations of language that sapphic artists are using to portray their experience. This study conducts this analysis with the knowledge that the methods of distant reading were not built to deal with these types of texts.[8] Urban Latin music song lyrics are slang-heavy, filled with non-standard manipulations of the Spanish language. They are repetitive, highly onomatopoeic, and at times play with the boundaries between Spanish, English, and Portuguese.[9] Therefore, this essay will not take distant reading as a direct path to knowledge-making but will instead engage in a two-way negotiation with its affordances, shortcomings, and future possibilities. In doing so, it points to the importance using of distant reading in conjunction with other interpretative methods, and the limitations of its "liberatory" corpus approach. In this case, distant reading aided in hypothesizing that sapphic artists simultaneously imitate and eschew the male-dominated language of desire in urban Latin music. This position is, in Elaine Showalter's terms, a rendezvous of the "muted zone" and the "wild zone" of women's writing, differentiated by their relationality to male convention.[10]

---

[3]Natalie M. Houston. "Distant Reading and Victorian Women's Poetry." In *The Cambridge Companion to Victorian Women's Poetry* (Cambridge University Press, 2019), 249–65.

[4]Joanna Byszuk, "The Voices of Doctor Who – How Stylometry Can Be Useful in Revealing New Information About TV Series," *Digital Humanities Quarterly* 014, no. 4 (2020).

[5]See for instance Elizabeth Monzingo and Daniel Shanahan. "The Expression of Self and Grief in the Nineteenth Century: An Analysis through Distant Readings." *Nineteenth-Century Music Review* 18, no. 1 (2021): 83–107.

[6]Henceforth, urban Latin music is used as an umbrella term for reggaetón, trap Latino, dembow, bachata, merengue, among other styles used by the artists in the corpus.

[7]In this essay, the use of the term "sapphic" is an intentional choice to encompass the variety of identities through which queer desire aimed at women presents itself in urban Latin music.

[8]Distant reading practices are based on dictionaries mostly compiled from newspaper sources, which contain fairly standard manifestations of language. See Hans Ole Hatzel, et al., "Machine Learning in Computational Literary Studies." *It - Information Technology* 65, no. 4–5 (2023), 201-202.

[9]Patricio Josué Velarde Dediós. "Reguetón de mujeres para mujeres: dos caminos de empoderamiento." *Antec: Revista Peruana de Investigación Musical* 5, no. 1 (2021): 91–109.

[10]Elaine Showalter, "Feminist Criticism in the Wilderness." *Critical Inquiry* 8, no. 2 (1981): 179–205.

2 Sapphic Desire in Urban Latin Music

In the last decade, women artists have slowly been carving out a space for themselves in urban Latin music. In a genre that blatantly advances men's utilization of women as media for interrelationality, and as sexual, often less-than-human objects, women artists have had to carefully tip-toe the conventions of the genre to cultivate their careers. In their earliest interventions, women were relegated to uncredited lyrical accompaniments and had to portray themselves as the objects of male desire.[11] As Velarde Dediós points out, this trend has shifted since the mid-2010s, with artists such as Becky G and Natti Natasha. This new wave is characterized by high levels of collaboration amongst women artists.

Most recently, a further development in this trend of women's participation has begun—the inclusion of sapphic artists. As the term suggests, this "sapphic" turn has not only signified the popularization of new, lesbian-identifying artists but also the release of songs about women-for-women desire by artists who usually write heterosexual lyrics, as was the case with Karol G and Kali Uchis' "Labios Mordidos." This development poses new theoretical challenges. How can women establish themselves as both subjects and objects of desire in a sphere built through the homosocial relations of men? To an extent, the answer to this question is experimentation. Reggaetón bears particular linguistic characteristics—from the hegemonic character of Puerto Rican slang to an established set of zoomorphic words to refer to women.[12] Part of the way sapphic artists are inserting themselves in this genre is by bending these conventions and imagining new possibilities for the lexical space of reggaetón.

Theorists of sapphic desire have previously encountered this very space of inventiveness. In "Refiguring Lesbian Desire," Elizabeth Grosz engages in an experimental reframing of lesbian desire away from male-centric, often phallocentric Freudian conceptions of "lack" and "absence." Grosz proposes a theory of desire borrowing from rhizomatic thinking, where it is defined in terms of "parts and bits and interconnections, [what] is new, [what] is exploratory, that opens up further spaces (…) and proliferates production."[13] In these terms, desire becomes a productive force, a force that does not incur hierarchy but rather flattens it. A force that inevitably engenders knowledge.

This productive space of sapphic desire is perhaps what Elaine Showalter was imagining when she theorized the "wild zone" of women's culture. This is the space in women's experience that is "no man's land." Because all spheres of society are mediated by male conventions, including language, the wild zone might even be a space of foreign tongues, a space outside the "prison-house" of language.[14] According to Showalter, the wild zone is a metaphorical proposition, because in reality most of women's writing is mediated through the "muted zone." In the muted zone, "muted groups must mediate their beliefs through the allowable forms of dominant structures."[15] Hence, Showalter proposes that we think of women's writing

---

[11]Velarde Dediós, 92.

[12]Pavlina Vankova. "Studying the Vocabulary of Reggaeton Song Lyrics." *Topics in Linguistics* 23, no. 2 (2022): 63–88; Angel Torres-Toukoumidis, et al. "Computational Analysis of Latin Music Songs Through Tokenization. Case of Female Artists and Reggaeton." In *Communication and Applied Technologies*, edited by Paulo Carlos López-López, et al., Smart Innovation, Systems and Technologies. (Singapore: Springer Nature, 2023), 527-35.

[13]Elizabeth Grosz. "Refiguring Lesbian Desire." In *Space, Time and Perversion*. (Routledge, 1995), 81.

[14]Showalter, 194.

[15]Showalter, 201.

as a double-voiced discourse, that toggles the line between the muted and wild zones, and that goes beyond the dualistic view that women writers either imitate their predecessors or revise them.[16]

In the "prison-house" of urban Latin lyrics, to express desire for a woman almost inevitably signifies her sexualization and reduction to non-human elements. This conundrum is palpable in many of the recent hits of sapphic artists.

To illustrate this, let's perform a brief close reading of the hit song "Riri" by artist Young Miko:

Le toco su cuerpo, eso es lo que ella ama,	I touch her body, that's what she loves,
y prende la webcam cuando está en su cama.	and she turns on her webcam when she's in bed
(…)	(…)
Ella se va viral en IG cuando publica	She goes viral on Instagram when she posts.
(…)	(…)
Pa' ver su contenido tienes que pagar el prime.	To watch her content, you must pay for a prime [subscription].[17]

To a listener familiar with urban Latin genres, these lyrics are reminiscent of a common trend in male-dominated songs—an overt discussion and desire for women engaging in sexual work. One such song is "Only Fans" by Young Martino, Myke Towers, and Lunay.

Con el tiempo vi un vídeo por ahí	In time, I saw a video laying around,
Y en OnlyFans rápido me suscribí	and quickly subscribed to her OnlyFans.
(…)	(…)
Y e' una santa en Instagram (así)	She's a saint on Instagram,
Y e' una puta en OnlyFans (¡ah!)	but a whore on OnlyFans.[18]

A close reader with little external knowledge would likely group "Riri" among women's songs that primarily imitate men's language of desire. However, in an interview for *Rolling Stone*, Young Miko asserted that her music— although "raunchy"— is meant for women and queer folks.[19] With this knowledge in mind, it is possible to revisit the passage with a new lens. Whereas "Only Fans" is largely about the male singer "I," "Riri" predominantly endows its female character with agency, speaking of what she loves, and only once using a reference to the "I" (even then, the pronoun is in its reflexive form 'me' + verb).

How many more of these delicate negotiations between the muted and wild zones would we encounter if we were able to look at a representative sample of sapphic urban Latin music? Is it possible to identify and potentially explain general trends in stylistic choices? This exercise will illustrate the areas of textual interpretation where distant reading can be helpful, confusing, or a steppingstone in a larger system of reading.

---

[16]Showalter, 204.

[17]Translated by the author.

[18]Translated by the author.

[19]Solá Santiago Frances, "Welcome to Young Miko's World," *Rolling Stone* (blog), June 26, 2023. https://www.rollingstone.com/music/music-features/young-miko-trap-kitty-bad-bunny-karol-g-1234771397/.

3 The Creation of a Corpus-Canon

The selection of a representative corpus is a key step in any distant reading analysis. This entity—the "corpus"—stems from the field of corpus linguistics, which aims "to construct and empirically analyze a sample of texts as the basis for the meaningful generalizations about a language or discourse domain."[20] An imperfect analogue in close-reading-based literary studies is the literary canon. Although a canon is constructed through a historical process of scholarly inquiry, rather than statistical sampling, it decidedly positions certain texts as "representative" of a movement or domain of literature. Natalie Houston states that computational feminist criticism allows for "explicit reflection on the decisions made in constructing the dataset."[21] For Houston, distant reading offers a departure away from literary canons that are often examined uncritically. Therefore, scholars of computational feminist criticism argue that a corpus, by making the boundaries of selection explicit, opens up new possibilities of analysis and representativeness that go beyond individual authors or literary canons.[22]

Nonetheless, the boundary between a corpus and a canon should be trodden carefully. As theorists of corpus linguistics have suggested, it is perhaps just as easy to introduce bias into a corpus as it is to drown in the incompleteness of a canon.[23] This realization was central in the process of creating a corpus for this study. Sapphic artists in the urban Latin genre are multifaceted—they differ in terms of age, nationality, sexual identity, and artistic goals. In other words, even though this domain of language is bound together by a shared identity, it seems to resist corpus-making, and by extension, canon-making. Appropriate selection criteria for sapphic urban Latin songs are hard to determine. To select artists solely based on popularity would be to conform to a canon advanced by mainstream music industries, the same canon that women's writing has for centuries attempted to resist. Therefore, the corpus-creation process for this study was iterative. It began with the charts and thus selected top songs by popular artists such as Young Miko, Karol G, and María Becerra. Then, I took the time to scrutinize other areas where emerging sapphic artists often promote their music. These included social media and Spotify playlists and led to the inclusion of smaller artists such as Chzter and Gudnana, as well as artists whose agenda is slightly more activism-based, such as Chocolate Remix.

Admittedly, corpus creation also inevitably incurred decisions on the exclusion of material. For instance, I decided to not include sapphic artists from Spain, except when in collaboration with Latin artists. Although several Spanish women singers are now ubiquitous in the Latin music mainstream, the historical development of Spanish women's music is markedly different from that of Latin America. Therefore, their inclusion would introduce noise and skew the corpus. The sampling strategy was also affected by the fact that most lyrics were scraped from Genius, and therefore limited by the songs available on the platform. However, an effort was made to manually retrieve lyrics to lesser-known songs from other online sites, such as AZlyrics. No songs were transcribed by ear, although this remains a viable route in future studies. Lastly, a sample of songs by men was also included in the corpus, to facilitate comparative analysis. This sample was stratified to match the nationalities of the women sub-corpus since different Latin American

---

[20]Jesse Egbert, et al. *Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness, (*Cambridge University Press, 2022): 5-6.

[21]Houston, 251.

[22]Houston, 255.

[23]Egbert, et al.

countries can bring drastically different vocabulary and slang to song-writing.[24] Likewise, the sub-corpus includes lesser-known male artists such as Dani Flow.

The resulting corpus contains 166 songs, of which 83 are predominantly written by sapphic women, and 83 are written by men.[25]
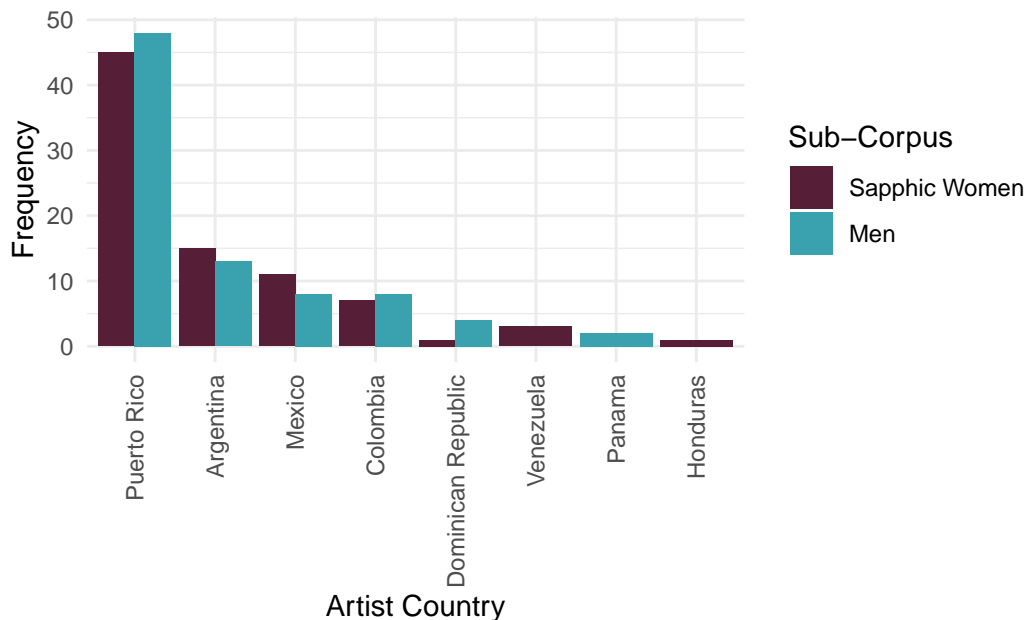
Fig. 1. Distribution of Songs in the Corpus by Country

However, this process still raised many questions. For instance, we may be able to determine whether a song is written by a woman, but how can we tell when a song is *about* a woman, and consequently, sapphic? Historically, sapphic singers have utilized different strategies to sing about women without making it apparent. A good example is the case of Ana Gabriel, who since the 1980s has been releasing songs about women, but masked behind male pronouns, such as "Simplemente Amigos" (Just Friends). Other songs forgo the gendered character of the Spanish language by addressing their lyrics to a "you," as is the case with Karol G's "CONTIGO." In this case, prudent decision-making was implemented. For instance, "CONTIGO" is accompanied by an explicitly sapphic music video and was therefore included in the corpus. Still, these complications of defining a "sapphic" song merit further discussion, and I encourage future researchers to delve further into this matter.

4 Results

The prospect of utilizing distant reading tools to understand the "female shift" in urban Latin music has interested previous researchers. In 2022, Angel Torres-Toukoumidis and his collaborators compiled a

---

[24]Vankova.

[25]It is important to mention that not all singers in the corpus were the only songwriters of their songs. However, the songs selected are ones that are evidently written predominantly by the (woman) performer.

corpus of 641 songs by 12 *reggaetoneras* and conducted a lexical analysis of their characteristics. Among their results, the authors found that women's reggaetón presented a high index of lexical diversity.[26] The authors interpreted this finding as counter-evidence to the popular belief that reggaetón is a genre with limited, simple vocabulary. Furthermore, the authors also found a lower usage of zoomorphic characteristics to self-designate, as is common in the genre's discussion of women. Although this study offers an essential introduction to the application of distant reading to urban Latin music, there are many ways in which its methods could be taken a step further.

My analysis aims to fill some of these methodological gaps and offer a view into the narrower population of sapphic urban Latin songs. In the first place, this study is based not only on a corpus of songs by women but also encompasses a comparative sub-corpus of songs by men. Secondly, this study will focus on a particular set of tokens that are commonly used in urban Latin music to denote women and desire.[27]

4.1. Frequencies

An appropriate first step when analyzing our corpus is to visualize a simple word cloud of the most common words (**Fig. 2a and b**). To make this visualization more helpful to our research goals, we deleted both Spanish and English stopwords.[28] Then, the wordclouds were separated by sub-corpus.

Fig. 2a. 100 Most Common Words in the Sapphic Women Sub–Corpus

---

[26]Torres-Toukoumidis, et al., 531.

[27]All analyses were done using R software, and a full **GitHub Repository** is available in Appendix A.

[28]This is due to the high usage of English loanwords in urban Latin music. Since this analysis uses `quanteda` tools in R, the stopwords deleted were those found in the `stopwords` package.

Fig. 2b. 100 Most Common Words in the Men Sub–Corpus

Even though the cleaning of stopwords brought forth more specific vocabulary, we can also see some of the challenges of using distant reading tools for urban music. These include onomatopeic tokens, such as "oh," "wuh," and "trr." Furthermore, many of the most frequent words in **Fig. 2a and b**, are contractions of stopwords that are commonly used in oral settings. These include "na," "pa," and "to'a" which stand for "nada (nothing)," "para (for)," and "toda (all-feminine)." Since stopword lists currently available for NLP (Natural Language Processing) are predominantly compiled using written language, the tool was unable to detect these contractions.

In much of urban music, it is common for artists to shout out their own names or names of collaborators/producers within their lyrics. In the lyrics-scraping process, I made a conscious decision to include these so-called "ad-libs," since according to previous literature, a lot of women's voices in urban music have historically been relegated to backing vocals.[29] However, these instances are visible in the above word clouds, as is the case with "Miko" (for Young Miko) and "Toki" (for Tokischa).

To experiment with these limitations, I manually compiled a list of tokens to remove from the analysis. These include common contractions of stopwords in Spanish, names of artists and producers, and onomatopeic tokens.[30]

Fig. 3. below shows a comparison of the most used tokens for both groups, before and after cleaning. In these plots, the x-axis was converted from token frequency to a measure of proportion calculated using $\frac{TokenFrequency}{TotalSubcorpusTokens}$.

---

[29] Dediós, 92.

[30] A full list of removed words can be found in Appendix B.

This is because, even though both sub-corpora contain the same number of songs, the men sub-corpus contained considerably more tokens. This raises interesting questions regarding possible differences in song length or singing speed between the two groups which merits further scholarly inquiry.
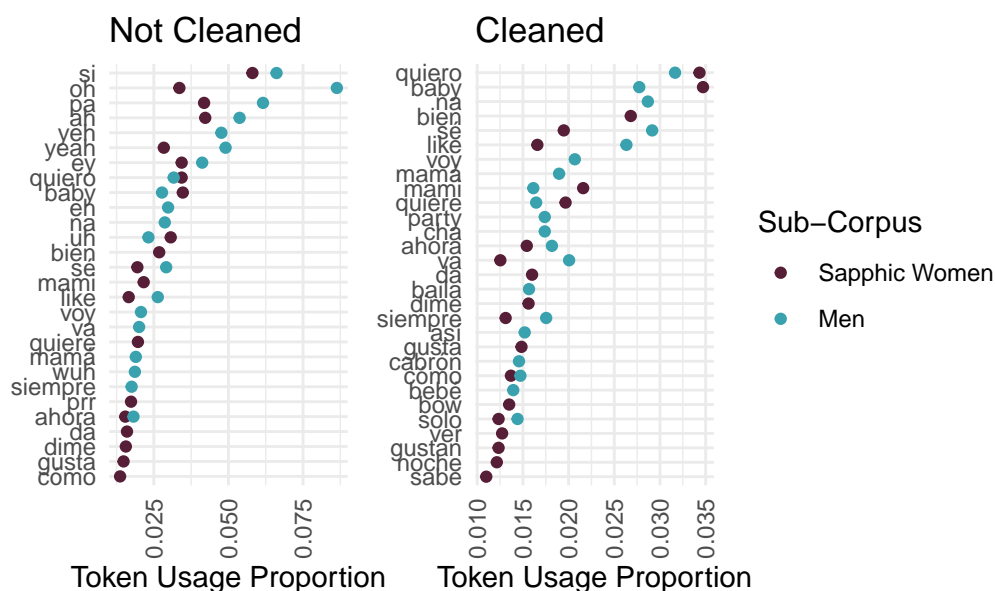

Fig. 3. Most Common Tokens by Sub–Corpus

In the uncleaned facet, we can observe that some of the onomatopoeic vocabulary seems to occupy a higher proportion of men's vocabulary in the corpus, like in the case of "oh" and "ah." However, other onomatopoeic tokens, such as "prr," are exclusively used by women. Words almost exclusively used by men include "party," "baila" (dance), and "cabrón" (Puerto Rican slang to denote 'asshole', gendered male). This could suggest that men are keener to situate their songs within party/club spaces. Words exclusively used by sapphic women include "bien" (good), "dime" (tell me), and "gusta" (he/she likes). These direct addresses of the object of the song, and a reference to the object's desires, could suggest that songs by women present a higher engagement with the thoughts of the women to which their songs are addressed. For instance, by quite literally asking them to speak.

Whether it is possible to compile a list of words to remove in a way that is objective and does not introduce bias into the dataset is worthy of debate. The following analyses will work with the entire corpus, including stopwords, since some of the words of interest include female pronouns and articles, such as "ella" (she/her).[31]

Table 1 below shows a number of these results where interesting comparisons were found across both groups. For instance, only the sapphic women sub-corpus uses the word "bestie." Perhaps this could be attributed to common storylines of friendship that exist in sapphic songs. Simultaneously, it was found that the men's sub-corpus uses the word "amiga" (female friend) at a higher proportion. This contradiction,

---

[31] For a full list of words investigated, see Appendix C.

9

and the meaning of these words in their specific context will be investigated further. Major differences were found for the words "bitch" and "booty." Along with "bestie," this could suggest a general trend for the usage of words in English to refer to women in sapphic songs, especially with regards to the body. As noted in **Fig. 3**, the word "gusta" (he/she likes) remains a notable difference between the two groups.

Table 1. Words of Interest Across both Groups (Selected)

	Token	Frequency	Rank	Song_Freq	Sub_Corpus	Proportion
5420	amiga	29	225	10	Men	0.0045
344	amiga	13	331	10	Sapphic Women	0.0025
5231	baby	177	45	56	Men	0.0277
27	baby	180	27	49	Sapphic Women	0.0347
114	bestie	43	112	3	Sapphic Women	0.0083
5857	bitch	9	646	6	Men	0.0014
85	bitch	57	84	16	Sapphic Women	0.0110
6541	booty	4	1285	3	Men	0.0006
136	booty	35	132	11	Sapphic Women	0.0067
5279	conmigo	80	92	29	Men	0.0125
89	conmigo	56	86	29	Sapphic Women	0.0108
5234	ella	160	48	38	Men	0.0251
34	ella	162	34	42	Sapphic Women	0.0312
5321	gusta	53	134	20	Men	0.0083
64	gusta	77	64	29	Sapphic Women	0.0148
5189	la	1252	3	83	Men	0.1962
4	la	845	4	81	Sapphic Women	0.1629

It is important to note, however, that at times these differences can be attributed to specific authors. For instance, the token "bitch" and its overrepresentation in the women's sub-corpus can likely be attributed to the artist Young Miko, as evidenced by **Fig. 4**. below.
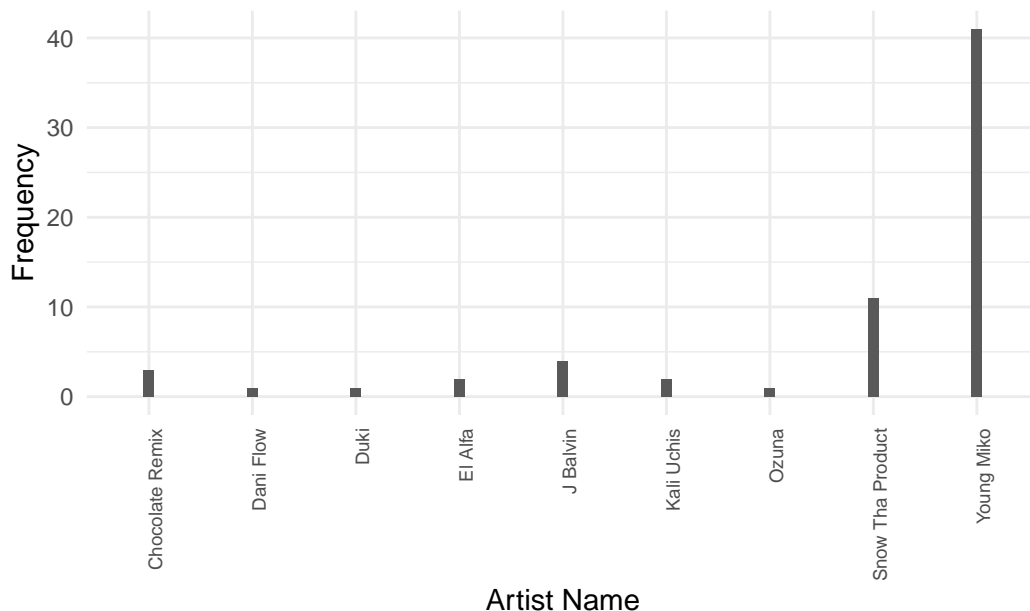
Fig. 4. Distribution of the Token 'Bitch' by Artist

The study of frequencies elucidates complications in the data-cleaning phase that any study of urban music should consider to avoid artificial conclusions. Furthermore, word frequencies suggested several hypotheses on stylistic differences, but the boundary between sapphic lived experiences and the artistic choices of individual authors must be examined critically.

4.2. Concordances

Another tool in distant reading that can provide further insight into our findings are concordances, which allow us to see the context in which a certain token appears in the corpus.

In **Table 2 and 3** below, I decided to look into one of the most sexualized terms to describe women in urban Latin music– "culo" (ass). In these contexts, we can see a certain level of lexical imitation from sapphic women. In many of these cases, the word "culo" is used in conjunction with reflexive pronouns such as "me," which emphasize the effect of an action on the speaker. In other words, both sub-corpora seem to refer to women's bodies in relation to the speaker's wants and desires. Furthermore, in the song "Deseándote" by Nath, the token is placed in close proximity to the car brand "Mercedes," engaging in a discourse of comparison of women with objects.

Table 2. Concordance of the Token 'Culo' in Men Sub-Corpus (Random Sample)

docname	pre	keyword	post	pattern
Soltera (Remix)_Daddy Yankee	Tito Hu huh A ese	culo	el traje le queda chiquito	culo*

docname	pre	keyword	post	pattern
La Mamá de la Mamá_El Alfa	Busca ese cuero que ese	culo	lo voy a hincha Le	culo*
Tirando Flow Sesh #11_Dani Flow	MOTOPAPI Brr tu chacalote Tu	culote	se ve ricote bien montadota	culo*
I Like It_J Balvin	hubo pues Como mueve el	culo	la dominicana Qué lo que	culo*
Banda de Camión (Remix)_El Alfa	gastando Sí Las babie su	culo	meneando Sí Y to lo	culo*

Table 3. Concordance of the Token 'Culo' in Sapphic Women Sub-Corpus (Selected)

docname	pre	keyword	post	pattern
Besties (Remix)_Young Miko	completa las modifico Qué Eso	culo	dan un cien lo certifico	culo*
Brinca_Young Miko	she's litty Uh Culo grande	culo	Iggy Rrr Que la busque	culo*
Deseándote_Nath	Pa qué Mercedes con ese	culote	Y cuando se suelta más	culo*
ID_Young Miko	Ja Perreando duro pégame ese	culo	Echa eso pa'cá canto e	culo*
Labios Mordidos_Kali Uchis & KAROL G	Dios me le bendiga ese	culo	que se pega uh uh	culo*

Concordances can also help us understand the role of the word "amiga" (female friend) in both groups. In the previous section, we observed how the sapphic sub-corpus exclusively uses the word "bestie" but includes a lower proportion of the word "amiga." **Table 4 and 5** reveal that in the men sub-corpus, the word "amiga" is used to refer to a woman's friends, towards whom the singer is hoping to make sexual advances. By contrast, in several instances of the sapphic sub-corpus, the word is used to describe an unsatisfactory friendship that the singer desires to be romantic. Such is the case in Gudnana's "AMG": "que ladilla aparentar que solo como amiga yo quiero tenerte" (how upsetting to pretend that I just want to have you as a friend.) This case could be interpreted as a reinvention of convention, or as a reflection of the specific lived experiences of sapphic women.

Table 4. Concordance of the Token 'Amiga' in Men Sub-Corpus (Random Sample)

docname	pre	keyword	post	pattern
No Fue (Remix)_Rauw Alejandro	Y girl dile a tu	amiga	que me haga coro Que	amiga*
Te Boté (Remix)_Bad Bunny	te mandé Y a tu	amiga	me clavé me la clavé	amiga*
No Fue (Remix)_Rauw Alejandro	parta Dice Dile a tu	amiga	que me haga coro Oh	amiga*

docname	pre	keyword	post	pattern
Tití Me Preguntó_Bad Bunny	Ey Hazle caso a tu	amiga	ella tiene razón Yo vo'a	amiga*
Baila Baila Baila (Remix)_Daddy Yankee	soltera Y taguea a su	amiga	pa romper la carretera Y	amiga*

Table 5. Concordance of the Token 'Amiga' in Sapphic Women Sub-Corpus (Selected)

docname	pre	keyword	post	pattern
AMG_Gudnana	ladilla aparentar que solo como	amiga	yo quero tenerte Aunque al	amiga*
Dime Como Hago_Maria Becerra	Y no quiero ser tu	amiga	Quiero ser la persona que	amiga*
Linda_Tokischa & Rosalía	estaba con La ROSALÍA Las	amigas	que se besan son la	amiga*

4.3 Keyness

One more test that we can run to look into the matter of lexical comparisons is a keyness test. This tool conducts a Chi-Squared statistical test to determine which words may distinguish a target group from a reference.[32]

In **Fig. 5**, the target group is the sapphic subcorpus, whereas the reference is the men's subcorpus. The tokens on the graph are those determined by the test to appear in the target and reference groups at a higher incidence than expected from chance. We can see some of the tokens we have already been investigating, such as "gustan," "bestie," and "bitch." However, many of the "distinctive" tokens are evidently associated with particular songs, such as "Chulo" and "Brinca," present in songs by the same name by Young Miko and collaborators. Therefore, it is valuable to question how much of these words' distinctiveness can be attributed to individual artists' choices in their personal projects, rather than their adherence to a general movement.

---

[32]For more on keyness see Lukas Sönning. "Evaluation of Keyness Metrics: Performance and Reliability." *Corpus Linguistics and Linguistic Theory* (2023).
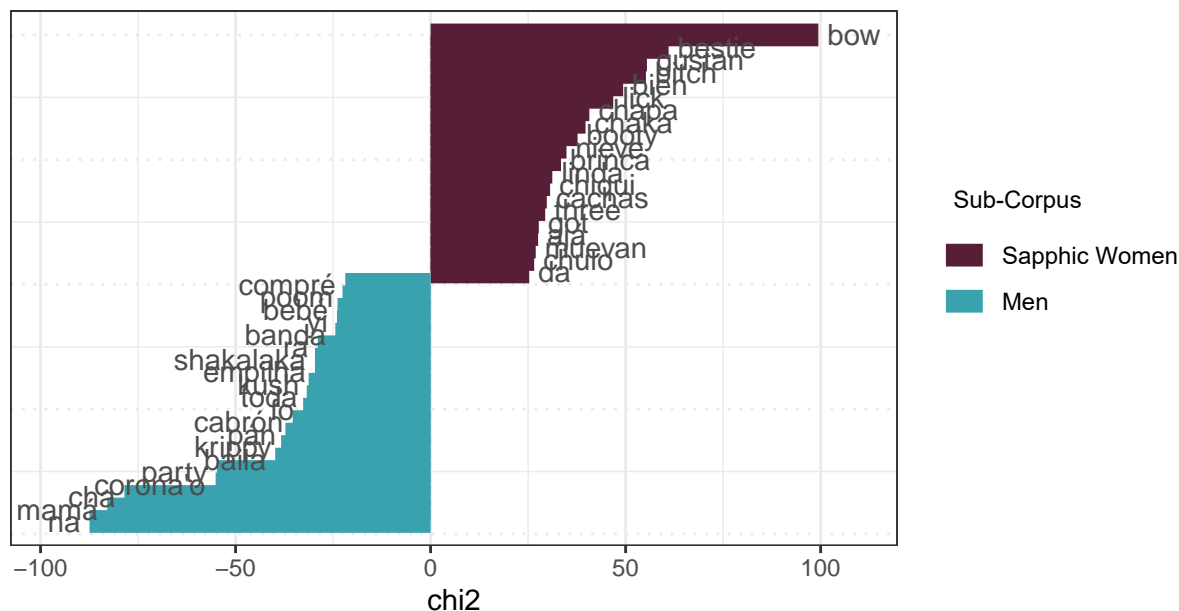
Fig. 5. Keyness Test Across Sub–Corpus

4.4 Stylometry

Lastly, since this study seeks to understand distinguishable differences in terms of vocabulary and style in the sapphic sub-corpus, stylometric analysis is an authoritative addition. Stylometric analysis is commonly used for tasks such as authorship attribution since it is capable of detecting patterns in sets of corpora, and consequently predicting the classification of new, unseen texts.[33] However, it is also helpful to visualize clusters of texts that are stylistically similar in terms of tokens used.[34]

I conducted a simple cluster analysis on the corpus, using 1000 tokens and the Eder's Delta distance measure, which has been shown by previous scholars to work best for languages with high rates of inflection.[35] Then, I labeled the results by Artist and by Sub-Corpus (using the color aesthetic).

In **Fig. 6** below, we can see that the Cluster Analysis was successful in grouping some of the artists' songs together. Many of Young Miko's songs (labeled in dark green) are plotted close to each other, especially those from her EP "Trap Kitty." The case was similar for Rauw Alejandro (labeled in blue), Paulo Londra (red), and Dani Flow (dark red).

In **Fig. 7**, the color aesthetic corresponds to the sub-corpus. Here, we can see a superposition of patterns. Firstly, we can see two main clusters of sapphic songs (in red), at opposite ends of the plot. The upper cluster largely includes songs by Young Miko, along with a few by Snow Tha Product. The lower cluster

---

[33] Eder, et al., 107.

[34] The analysis was done using the R package `stylo`, which allows for an intuitive introduction to the method.

[35] José Calvo Tello. "Entendiendo Delta desde las humanidades." *Caracteres: estudios culturales y críticos de la esfera digital* 5, no. 1 (2016): 140–76.
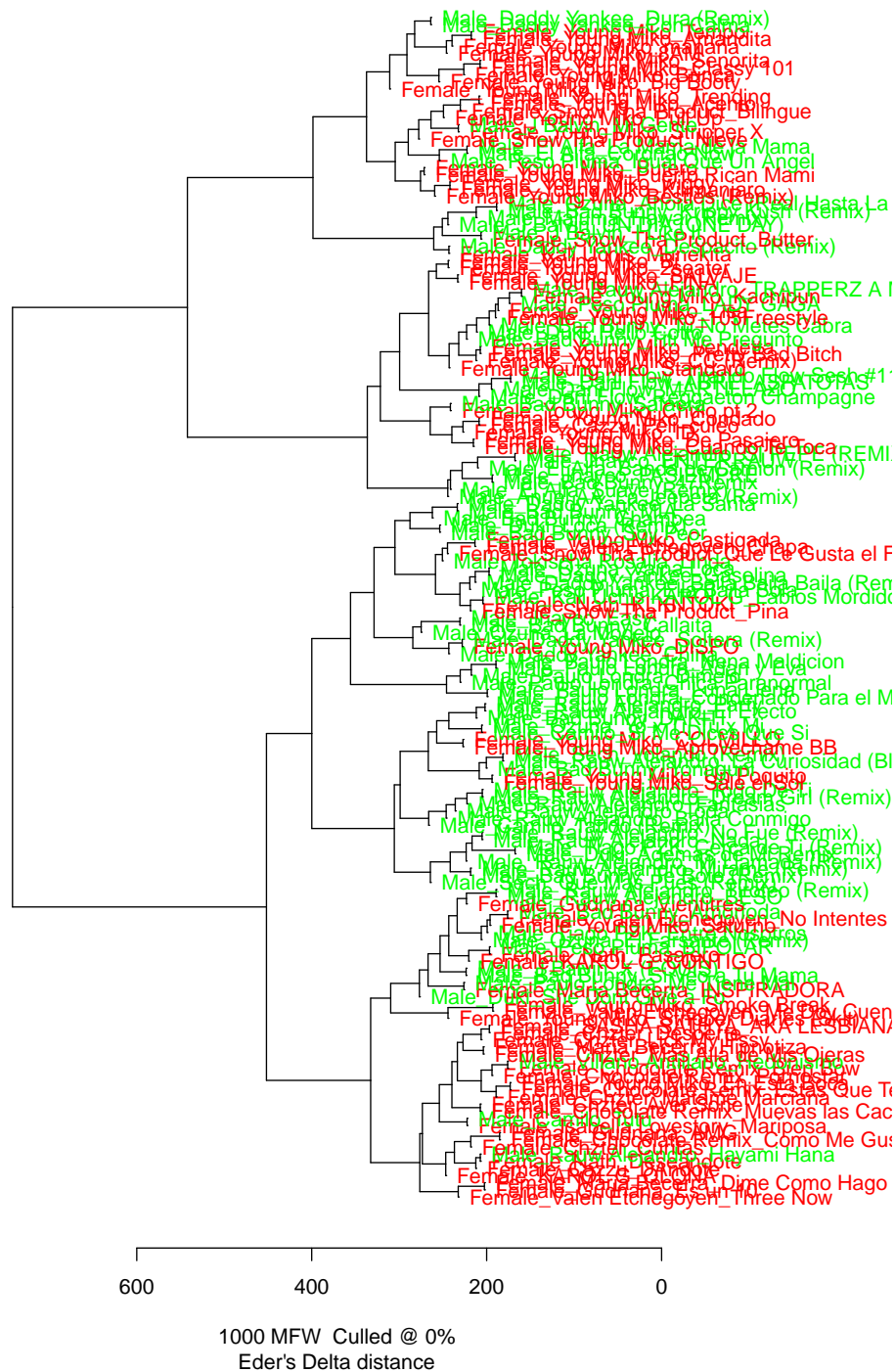
shows songs by Chzter, Gudnana, Maria Becerra, among others. Although there are distinguishable clusters of sapphic songs, many others are also interspersed throughout the plot with male-authored songs. We may imagine these intersecting clusters as what Showalter described as a double-voiced discourse. Although in some areas of this creative space, women's writing seems to carve its own pocket of invention, in other areas, it is intermixed with its conventional surroundings.

Fig. 6. Grouped by Artist
Cluster Analysis

1000 MFW Culled @ 0%
Eder's Delta distance
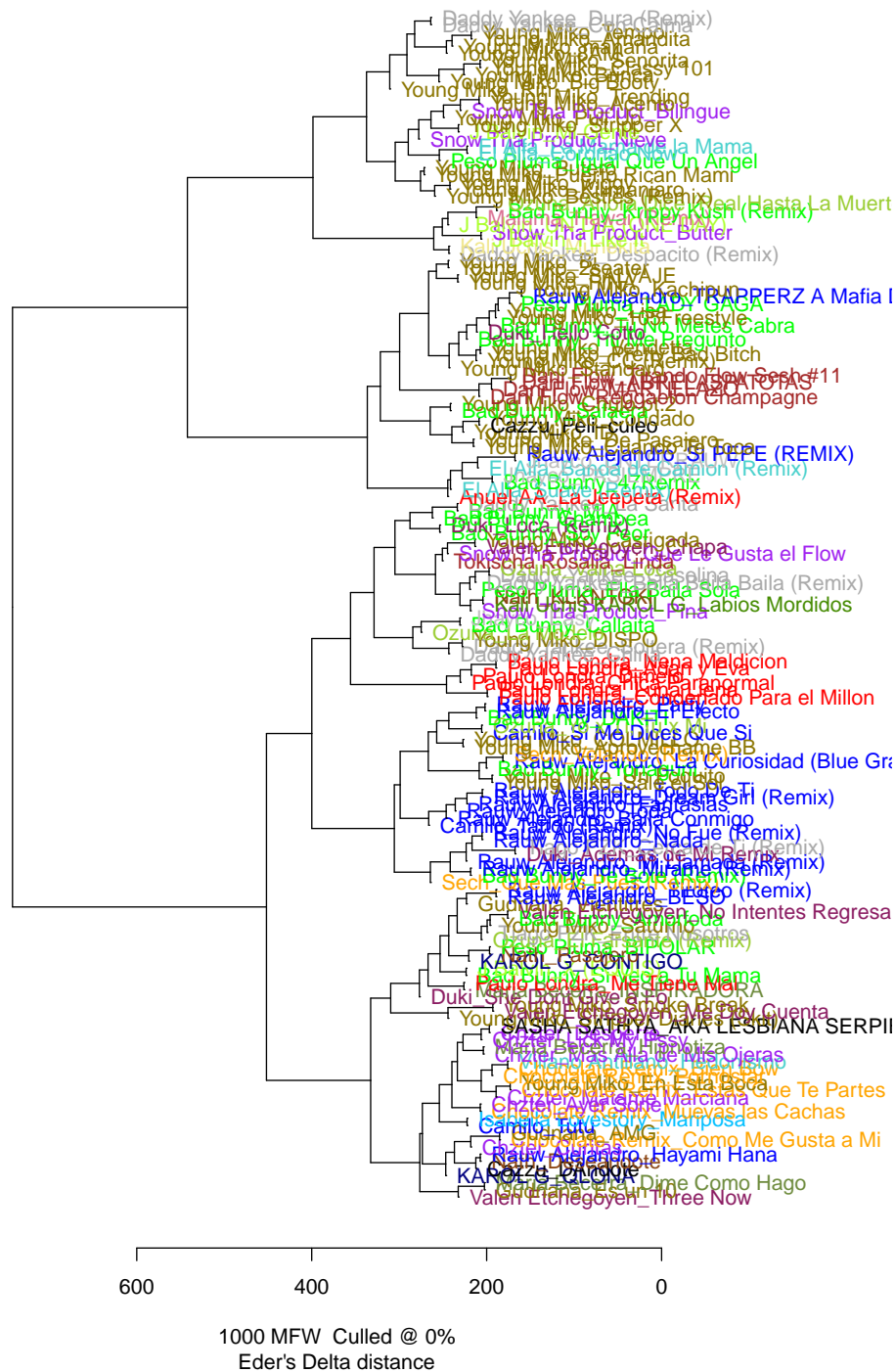
Fig. 7. Grouped by Sub−Corpus
Cluster Analysis

1000 MFW Culled @ 0%
Eder's Delta distance

17

5 Conclusion

At the beginning of this essay, we identified two facets of distant reading that are commonly discussed among scholars. The first, its genealogical relationship with close reading. Secondly, distant reading's perceived ability to expand literary inquiry beyond an established canon through self-reflective data collection and analysis practices. The study of sapphic urban Latin music elucidated the push-and-pull between the affordances and shortcomings of distant reading methods. In some ways, our close reading of "Riri" was not fundamentally different from our analysis of the corpus. In both cases, interpretation hinged on specific words, their lexical context, and their position within a sentence. There are many avenues towards the interpretation of a text, and perhaps often the best approach is an omnivorous one.

For instance, take one of the insights that we derived from our distant reading– the difference in the usage of the word "amiga" (female friend) among the two sub-corpora. There are a couple of ways we could have arrived at this observation. For instance, we could have decided to do a close reading of "Linda" by Tokischa & Rosalía, a widely popular song, and its discussion of homoerotic friendships. We could have taken one of Gudnana's social media videos, where she speaks about the story behind her song "AMG," which describes an unrequited crush on a friend. Each of these approaches would have come with specific affordances. The distant reading approach allowed us to take in a relatively large number of texts and be able to draw general patterns across them. It does not, however, provide the amount of granularity that other approaches can. For instance, distant reading easily confounds tokens found repeatedly in a specific text for tokens that are generally representative of a domain.

If handled without care, distant reading does exactly what it seeks to avoid– it draws generalizations and canonizes a corpus. Without proper knowledge of the domain, as well as an elementary understanding of the mechanisms behind the method, distant reading can even provide misleading insights. If, as posited by Houston, computational feminist literary criticism is to break free from the patriarchal structures of canon-making, it must engage face-to-face with the statistical impulse that lies at its core– to generalize. Limitations of representativeness should therefore be taken into honest account when assessing the validity of analytic claims. And, in turn, distant reading insights should rarely stand on their own, instead of complementing existing or opening up new instances of inquiry. Just like the bounds of the written will never completely break the prison-house of language, distant reading cannot on its own free us from the prison-house of canon-making.

6 Appendix

6.1 Appendix A

A full **GitHub Repository** is available at this link.

6.2 Appendix B

List of removed words, also available as `removed_words.txt` on the GitHub repository.

uh
ah
yeah
miko
yankee
bunny
ey
wuh
woh
oh
eh
toki
caleb
pa
si
yeh
ay
to
yah
mmm
prr
trr
huh
uah
to'a
might
rauw
bam
bum

6.3 Appendix C

List of words of interest, also available as `words_of_interest.txt` on the GitHub repository.

la
las
una
baby
ella
you
bien
mami
esta

bitch
gusta
conmigo
eres
culo
dale
bestie
esa
contigo
duro
ma
bad
fuck
she
booty
puta
loca
chapa
nena
gata
cama
linda
bebé
mala
girl
cuerpo
muñequita
puti
shorty
gyal
niña
mía
amiga
mujer

7 Bibliography

Acosta-Aguilar, Karina. "How Reggaeton Is Enabling Self-Expression, Feminism, and Dismantling Machismo." *Medium* (blog), May 4, 2023. https://medium.com/@karinaacost7/how-reggaeton-is-enabling-self-expression-feminism-and-dismantling-machismo-6462b1a7e588.

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly (Seattle)* 78, no. 1 (2017): 77–106. https://doi.org/10.1

215/00267929-3699787.

Byszuk, Joanna. "The Voices of Doctor Who – How Stylometry Can Be Useful in Revealing New Information About TV Series." *Digital Humanities Quarterly* 014, no. 4 (December 20, 2020).

Calvo Tello, José. "Entendiendo Delta desde las humanidades." *Caracteres: estudios culturales y críticos de la esfera digital* 5, no. 1 (2016): 140–76.

Dediós, Patricio Josué Velarde. "Reguetón de mujeres para mujeres: dos caminos de empoderamiento." *Antec: Revista Peruana de Investigación Musical* 5, no. 1 (August 14, 2021): 91–109. https://doi.org/10.62230/antec.v5i1.112.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8 (January 1, 2016): 107–21. https://doi.org/10.32614/RJ-2016-007.

Egbert, Jesse, Douglas Biber, and Bethany Gray. *Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness*. Cambridge University Press, 2022. https://doi.org/10.1017/9781316584880.

Goldman, Dara E. "Walk like a Woman, Talk like a Man: Ivy Queen's Troubling of Gender." *Latino Studies* 15, no. 4 (November 1, 2017): 439–57. https://doi.org/10.1057/s41276-017-0088-5.

Grosz, Elizabeth. "Refiguring Lesbian Desire." In *Space, Time and Perversion*. Routledge, 1995.

Hatzel, Hans Ole, Haimo Stiemer, Chris Biemann, and Evelyn Gius. "Machine Learning in Computational Literary Studies." *It - Information Technology* 65, no. 4–5 (August 1, 2023): 200–217. https://doi.org/10.1515/itit-2023-0041.

Houston, Natalie M. "Distant Reading and Victorian Women's Poetry." In *The Cambridge Companion to Victorian Women's Poetry*, 249–65. Cambridge University Press, 2019. https://doi.org/10.1017/9781316856543.017.

LeBrón, Marisol. "'Con Un Flow Natural': Sonic Affinities and Reggaeton Nationalism." *Women & Performance: A Journal of Feminist Theory* 21, no. 2 (July 1, 2011): 219–33. https://doi.org/10.1080/0740770X.2011.607598.

Monzingo, Elizabeth, and Daniel Shanahan. "The Expression of Self and Grief in the Nineteenth Century: An Analysis through Distant Readings." *Nineteenth-Century Music Review* 18, no. 1 (2021): 83–107. https://doi.org/10.1017/S1479409819000697.

Mulvey, Laura. "Visual Pleasure and Narrative Cinema." In *Feminism and Film Theory*, 57–68, 2013. https://doi.org/10.4324/9780203699362.

Noriega, Dulce Asela Martínez. "Música, imagen y sexualidad: el reggaetón y las asimetrías de género." *El Cotidiano*, no. 186 (2014): 63–67.

Rivera-Servera, Ramón H. *Performing Queer Latinidad: Dance, Sexuality, Politics*. 1st ed. Triangulations: Lesbian/Gay/Queer Theater/Drama/Performance. Ann Arbor: University of Michigan Press, 2012. https://doi.org/10.3998/mpub.2395967.

Showalter, Elaine. "Feminist Criticism in the Wilderness." *Critical Inquiry* 8, no. 2 (1981): 179–205.

Solá-Santiago, Frances. "Welcome to Young Miko's World." *Rolling Stone* (blog), June 26, 2023. https://www.rollingstone.com/music/music-features/young-miko-trap-kitty-bad-bunny-karol-g-1234771397/.

Sönning, Lukas. "Evaluation of Keyness Metrics: Performance and Reliability." *Corpus Linguistics and Linguistic Theory*, April 27, 2023. https://doi.org/10.1515/cllt-2022-0116.

Toro, Ximena de. "Métele con candela pa' que todas las gatas se muevan. Identidades de género, cuerpo y sexualidad en el reggaetón." *Revista Punto Género*, no. 1 (October 13, 2011). https://doi.org/10.5354/2735-7473.2011.16824.

Torres-Toukoumidis, Angel, Cristian Orlando Coronel Quezada, Jenny Pontón, and Isidro Marín-Gutiérrez. "Computational Analysis of Latin Music Songs Through Tokenization. Case of Female Artists and Reggaeton." In *Communication and Applied Technologies*, edited by Paulo Carlos López-López, Daniel Barredo, Ángel Torres-Toukoumidis, Andrea De-Santis, and Óscar Avilés, 527–35. Smart Innovation, Systems and Technologies. Singapore: Springer Nature, 2023. https://doi.org/10.1007/978-981-19-6347-6_47.

Vankova, Pavlina. "Studying the Vocabulary of Reggaeton Song Lyrics." *Topics in Linguistics* 23, no. 2 (2022): 63–88. https://doi.org/10.2478/topling-2022-0012.