

Lexico-visual Analysis and Lexical Frequencies in Urban Music: a Methodological Proposal for the Study of Audiovisual Narratives

Análisis lírico-visual y frecuencias léxicas en la música urbana: una propuesta metodológica para el estudio de narrativas audiovisuales

Paz Crisóstomo Flores. Universidad Finis Terrae (Chile)

Academic and researcher at the School of Journalism at Finis Terrae University and coordinator of the Analysis and Research Department. PhD in Media Research (UC3M). She has developed projects and publications in leading magazines. Areas of expertise: research methodology, journalistic quality, news coverage, television, and unforeseen events.

ORCID: <https://orcid.org/0000-0002-9761-2112>

Nicolás Hernández Pinto. Universidad Finis Terrae (Chile)

Journalist and graduate in Social Communication with a minor in Strategic Communication from Finis Terrae University. Winner of the 2022 Academic Excellence Award and the Human Rights Memory Journalism Award from the Museum of Memory, he covered the 2023 Pan American Games in Santiago and has studied Language Pedagogy. ORCID: <https://orcid.org/0009-0004-7741-4575>

Vicente Higuera Cortés. Universidad Finis Terrae (Chile)

Journalist and graduate in Social Communication with a minor in Strategic Communication from Finis Terrae University. Winner of the Human Rights Memory Journalism Award from the Museum of Memory, he covered the 2023 Pan American Games in Santiago and is the founder of Engira.cl, an independent media outlet that covers Chilean urban music news on social media. ORCID: <https://orcid.org/0009-0003-3784-0572>

Article received: 18/10/2025 – Accepted: 21/11/2025

Abstract:

This article introduces a lexical-visual model for multimodal analysis of audiovisual narratives, combining a lexical dictionary, a visual coding matrix and a relational component based on network analysis. The proposal integrates lexical counts, visual codification and inter-coder reliability assessment through Cohen's Kappa, enabling the operationalization of categories associated with representations of status, risk and transgression. The model is validated through a pilot application in a small corpus of Chilean urban music, used solely to demonstrate the instrument's operability. Triangulation between lexical and visual dimensions and the computation of a simple lexical/visual convergence index allow for evaluating the model's internal coherence

and analytical consistency. The article discusses the methodological scope, limitations and potential transferability of the model to other audiovisual contexts.

Keywords:

Lexical-visual model; Multimodal analysis; Audiovisual content analysis; Inter-coder reliability; Chilean urban music.

Resumen:

Este artículo presenta un modelo léxico-visual para el análisis multimodal de narrativas audiovisuales, integrado por un diccionario léxico, una matriz visual de codificación y una dimensión relacional basada en análisis de redes. La propuesta combina conteo léxico, codificación visual y evaluación de fiabilidad mediante el coeficiente Kappa de Cohen, permitiendo operacionalizar categorías asociadas a representaciones de estatus, riesgo y transgresión. El modelo se valida a través de una aplicación piloto en un corpus acotado de música urbana chilena, utilizada únicamente como demostración de su operatividad. La triangulación entre dimensiones léxicas y visuales y el cálculo de un índice de convergencia visual/léxico permiten evaluar la coherencia interna del instrumento y su potencial para investigaciones futuras. Se discuten los alcances, limitaciones y posibilidades de transferencia del modelo a otros contextos audiovisuales.

Palabras clave:

Modelo léxico-visual; Análisis multimodal; Análisis de contenido audiovisual; Fiabilidad intercodificador; Música urbana chilena.

1. Introduction

Since the emergence of Chilean trap on international stages, urban music has come to occupy a central place in public debates surrounding youth, culture, and security. Its narrative—combining global rhythms with local storytelling—incorporates themes associated with drug consumption and trafficking, the carrying of weapons, theft, and material ostentation (Díez, 2020). In parallel, several artists have faced judicial proceedings that have amplified the media visibility of the phenomenon, reinforcing the association between urban music and crime (Ruiz, 2022). This coverage, often focused on police spectacle rather than on social complexity, has contributed to the consolidation of a simplified and moralizing interpretation of the genre (Deflem, 2020), underscoring the need to address it through more systematic analytical tools.

In this context, media discourses surrounding Chilean urban music tend to reproduce stigmas associated with working-class youth, reinforcing the notion that cultural production operates as a space of risk or deviance. However, the music videos and lyrics within this repertoire also constitute aesthetic and political narratives that resignify inequality and social hierarchies. The ostentation of goods, weapons, or luxury brands can be interpreted not only as symbols of transgression, but also as codes of status, belonging, and cultural resistance. It is precisely this ambivalence that makes it

necessary to develop methodological approaches capable of translating these repertoires into observable and comparable categories across different audiovisual products.

Latin American scholarship on narcoculture provides a useful conceptual framework for this analysis. Mondaca (2012) defines it as a cultural process that integrates symbols, norms, and worldviews surrounding the illegal economy, projecting values and lifestyles beyond drug trafficking itself. From this perspective, *narco-aesthetics* emerges as a practice of visual and discursive ostentation in which luxury, violence, and masculinity are articulated around the promise of success (Rincón, 2009; Sandoval, 2020). In Chile, this aesthetic is recontextualized within urban neighborhoods shaped by tensions between social exclusion and aspirational consumption (Vásquez, 2017), providing a particularly fertile repertoire of signs to be operationalized into lexical and visual categories within a multimodal analytical model.

This study approaches the glorification of crime as a communicational phenomenon that manifests simultaneously across two analytical planes: the lyrical plane (words, phrases, and semantic fields that refer to criminal or narcocultural practices) and the visual plane (scenes, objects, attitudes, and symbols that represent them). Its primary objective is to propose and validate a methodological model capable of examining the convergence between these two planes through the construction of a lexical dictionary and a visual coding matrix, integrating quantitative analytical techniques (lexical frequencies, intercoder reliability) with qualitative approaches (semiotic and contextual analysis). The article therefore focuses on the formulation and evaluation of this model, rather than on the substantive interpretation of the corpus employed as a pilot application.

In this way, the article seeks to contribute to the field of communication studies by offering a replicable instrument that combines the systematic rigor of lexical counting with the interpretative sensitivity of audiovisual analysis. The proposal is illustrated through a pilot application to a corpus of Chilean urban music videos and lyrics, used solely as a testing ground for the model. The text is organized into three sections: (i) conceptual and methodological foundations for the operationalization of lexical and visual categories; (ii) model design and instrument construction; and (iii) model validation based on indicators of reliability and lexical–visual convergence. Finally, the article discusses the scope, limitations, and potential for transferring the methodology to other contexts and objects of study.

1.1. The glorification of crime as a communicational object

The concept of the glorification of crime has traditionally been defined within the legal field as the public exaltation or justification of acts classified as criminal. In Chile, it is regulated by Law No. 12,927 on State Security and Law No. 20,000 on drugs, which penalize forms of incitement or legitimization of criminal activity. However, this sanctioning framework is insufficient to fully grasp the phenomenon from a communicational and cultural perspective. Within contemporary media languages, glorification does not operate as mere direct incitement, but rather as a narrativization of transgression: a repertoire of symbols, gestures, and narratives that express experiences of inequality, precarity, and resistance through aesthetic and performative resources, which can be observed and coded across lexical and visual dimensions.

As Martín-Barbero (2003) argues, popular languages articulate situated forms of knowledge in which youth cultures transform exclusion into expressivity. From a similar perspective, Reguillo (2000) observes that Latin American urban youth rewrite risk and fear in aesthetic terms, turning violence into a form of visibility and belonging. Thus, what is interpreted as glorification within a legal register can be understood, from a communicational standpoint, as the staging of the social limit: a symbolic dramatization of power, desire, and exclusion that materializes through specific words, images, and performances.

From this perspective, Chilean urban music functions as a discursive device in which transgression is aestheticized and encoded. Following Hall (1997), all representation is productive: the media do not merely reflect reality but discursively construct it. Within this framework, references to weapons, drugs, or criminal acts do not necessarily imply promotion or literal glorification, but rather the configuration of meaning within a symbolic field in which marginality is reinterpreted as cultural value and can be analyzed through specific content categories.

This process can also be linked to the Foucauldian notion of the “discursive dispositif,” through which society produces, regulates, and recontextualizes that which it declares prohibited: the visibility of crime becomes a field of enunciation in which subaltern identities claim a place through the language of transgression. Authors such as Wacquant (2007) and Hartley (2010) offer a complementary reading by showing that, in contexts of structural inequality, crime ceases to function solely as an illegal act and instead becomes a symbolic form of resistance and recognition in scenarios where visibility and consumption operate as new forms of symbolic citizenship.

Communicational analysis shifts the focus away from the literal interpretation of the “dangerous” message toward the social function of discourse: how artistic expressions produce meaning, contest stigmas, and reconfigure the boundaries between what is permitted and what is transgressive. This interpretative turn makes it possible to operationalize the glorification of crime into observable dimensions—lexical and visual—that do not pass moral judgment on content, but instead measure its frequency, recurrence, and narrative function within the contemporary media ecosystem. On this basis, the categories of the lexical dictionary and the visual matrix of the proposed model are constructed, translating these notions into codifiable and replicable variables for the analysis of audiovisual narratives.

1.2. Narcoculture and Narco-Aesthetics: between symbolic value and representation

As Mondaca (2012) argues, the effectiveness of narcoculture does not lie solely in narrating crime, but in organizing desires, fears, and status through aesthetic repertoires that circulate across music, fashion, architecture, and audiovisual narratives. In this way, it projects models of success anchored in the logic of risk, rapid gain, and public visibility. Within this framework, *narco-aesthetics* refers to the visual-material dimension of this repertoire: an aesthetics of excess that translates accumulation—licit or illicit—into visible signs of power. Rincón (2009) describes it as a culture of ostentation (“cash, weapons, women, loud music, flashy cars”), while Sandoval (2020)

shows that these aesthetic challenges “legitimate” criteria of taste by proposing an aesthetic of having “made it”.

In Chile, according to Vásquez (2017), the country “imprints its own mark on the narco” by hybridizing urban inequality, aspirational consumption, and neighborhood-based forms of belonging. This specificity turns *narco-aesthetics* into a grammar of visibility: a way of saying “we are here” from historically stigmatized territories, where brands, jewelry, cars, and corporeal displays function as insignia of recognition. This phenomenon is further intensified on digital platforms and in music videos, where—following the logic of cultural mediatization—showing is as significant as saying, and elements such as vehicles, money, clothing, and bodily poses stabilize as recognizable repertoires that can be coded into visual categories.

However, not everything constitutes glorification: *narco-aesthetics* coexists with ambiguities (irony, fiction, stylization) and tensions (class, gender, race) that complicate its interpretation. In many cases, violence is aestheticized—shifted from the act to style—and luxury operates as a metonymy of power rather than as evidence of criminal activity. This complexity makes it necessary to explicitly define the criteria used to determine which signs are coded and how their function within the audiovisual narrative is interpreted.

For the purposes of the instrument, these notions are rendered observable and codifiable across two analytical planes that are subsequently triangulated: (1) the lyrical (lexical) plane, through semantic fields associated with weapons, drugs, crime, and luxury, operationalized in a dictionary that incorporates lemmas, local variants, anglicisms, and compound expressions; and (2) the visual (image) plane, through objects, settings, attitudes, clothing, vehicles, and money as material indicators of status and power, accompanied by decision rules regarding narrative centrality and degrees of predominance (high, medium, low).

In this way, narcoculture ceases to be an abstraction and becomes a measurable system of signs, suitable for lexical counting and semiotic coding. This makes it possible to assess convergences (what is said and what is shown) and dissonances (what is said without being shown, or vice versa) not only in Chilean urban music, but also in other audiovisual products in which imaginaries of risk, consumption, and power circulate.

1.3. Lyrical–Visual Narrative: from discourse to performativity

Urban genre productions are multimodal texts in which lyrics, voice, corporeality, and image compose an interdependent semiotic dispositif (Kress & Van Leeuwen, 2001). The music video does not mechanically “illustrate” the lyrics; rather, it expands their meaning into the realms of desire, the body, and performativity, generating emphases, ellipses, and redundancies that reconfigure the narrative. In the era of mediatization, cultural discourses are organized according to logics of platforms and visibility: the artist becomes a medium in and of themselves, optimizing signs and styles to circulate and connect with audiences (Hjarvard, 2015). This dynamic reinforces the need for instruments capable of systematically capturing the articulation between word, image, and performance.

From the perspective of cultural semiotics, the music video is a syncretic text that articulates verbal, visual, and sonic systems within a single communicative act (Lotman, 1996). Its analysis requires observing both convergence (reinforcement) and divergence (displacement or contradiction) between what is said (lexical) and what is shown (visual). In operational terms, this entails distinguishing among: (1) narrative centrality (close-ups, focalization, repetition) versus setting or ambience (peripheral presences); (2) performative indices (gestures, poses, direct gaze to the camera) as opposed to object-based indicators (weapons, money, cars); and (3) diegetic correspondence (when the lyrics enunciate what the image performs) versus misalignment (when the image stylizes or neutralizes what is enunciated). These distinctions are directly translated into observable variables within the lexical–visual model.

On this basis, the study proposes an integrated methodology that combines lexical counting and semiotic reading in order to identify patterns of recurrence between semantic fields (crime, luxury, violence, drugs) and their visual equivalents (weapons, money, vehicles, gestures of power). Operationally, (1) at the lyrical level, the dictionary groups lemmas and colloquial variants by semantic field and enables the measurement of frequency and distribution; and (2) at the visual level, the codebook records the presence and degree of predominance (high, medium, low) of objects, settings, and attitudes, incorporating decision rules for borderline cases (e.g., ambiguous props, symbols, irony), so that the relationship between lyrics and image can be analyzed using comparable and replicable criteria.

The dual (lexical–visual) reading makes it possible to quantify repetition (frequencies, proportions) and to qualitatively assess its narrative function (reinforcement, displacement, contradiction). In this way, it becomes possible to identify when the visual dimension amplifies what is enunciated in the lyrics (convergence) and when it attenuates or repositions it through processes of aestheticization (dissonance). Beyond the specific case, the proposed multimodal approach allows for a shift from an inventory of “themes” to a cartography of discursive operations, offering a procedure applicable to diverse audiovisual corpora in which semantic fields and visual repertoires are articulated under logics of mediation and performativity.

1.4. *Networks of collaboration and symbolic circulation*

Chilean urban music is sustained by networks of collaboration that operate simultaneously as market strategies and regimes of symbolic legitimation. As Riquelme et al. (2022) note, these alliances generate spaces of cooperation and belonging in which featuring practices expand reach and consolidate communities of style. Within the logic described by Olivares (2023), digital circulation (playlists, featurings, collective labels) determines visibility and positioning, shifting value from the isolated work to the web of relationships that sustains it. This dynamic makes it pertinent to incorporate a relational dimension into the analysis of these products.

Within the framework of *narco-aesthetics*, networks do not merely distribute music; they disseminate shared imaginaries. The same repertoires of luxury, risk, or violence are replicated through processes of symbolic contagion among artists, configuring a symbolic ecology in which signs and poses circulate and stabilize as common codes.

This relational web is therefore not confined to the economic or industrial plane, but can also be described and analyzed through the links between actors and the thematic co-occurrences that structure their productions.

For the purposes of this study, network analysis is incorporated as a relational dimension complementary to the lexical–visual component:

- (1) Nodes: artists included in the corpus.
- (2) Edges: credited collaborations (A feat. B) in the analyzed tracks.
- (3) Edge weight: number of shared collaborations (or co-occurrences during the 2019–2024 period).

Optionally, a thematic projection can be incorporated through an artist-by-category matrix (luxury, weapons, drugs, crime) to estimate semantic affinities and to observe whether collaborations co-vary with specific semantic fields.

On this basis, structural indicators are calculated to account for symbolic circulation:

- (1) Degree centrality: relational visibility (who collaborates the most).
- (2) Betweenness centrality: bridging roles that connect sub-scenes and diffuse codes across clusters.
- (3) Modularity / communities: identification of stylistic blocks (micro-scenes) characterized by similar visual and lexical repertoires.
- (4) Density and diameter: overall scene cohesion and the potential reach of symbolic contagion.

Visually, the collaboration graph makes it possible to map nodes (artists) and edges (ties), and to overlay semantic attributes (for example, color by dominant category or size by centrality) to observe how specific symbolic repertoires propagate through creative and commercial ties.

This approach extends textual analysis toward a cartography of symbolic power: it reveals which actors anchor dominant codes, which mediate between communities, and in which areas of the network lexical–visual convergence is concentrated. In sum, the relational dimension of the model helps explain why certain signs become consolidated—not merely which signs appear—and offers a transferable framework for studying the diffusion of imaginaries across different audiovisual scenes, beyond the specific case of Chilean urban music.

2. Lexical–visual model: a methodological proposal

This section presents the lexical–visual model developed for the multimodal analysis of audiovisual narratives. Rather than describing a methodology applied to a specific case, its purpose is to introduce a replicable instrument composed of a lexical dictionary, a visual matrix, and a relational component based on network analysis. Its application to a corpus of Chilean urban music is understood as a pilot test aimed at validating the model’s operational clarity, reliability, and internal coherence.

2.1. Research objectives

2.1.1. General objective

To design a mixed-methods methodological model for the multimodal analysis of audiovisual narratives, integrating a lexical dictionary, a visual coding matrix, and a relational dimension oriented toward the study of collaboration networks.

2.1.2. Specific objectives

SO1. To construct a lexical dictionary that operationalizes semantic fields associated with representations of crime, status, luxury, and violence.

SO2. To apply a visual matrix that enables the identification of relevant objects, settings, attitudes, and symbols within audiovisual narratives.

SO3. To assess the model's intercoder reliability using Cohen's Kappa coefficient and successive refinements of the codebook.

SO4. To apply the model to a pilot corpus of Chilean urban music to demonstrate its operational viability and analytical consistency.

2.2. Research design and general approach

This study is grounded in a mixed-methods design of multimodal content analysis, combining quantitative strategies, lexical counting and frequency measures—with a qualitative semiotic and discursive approach. It adopts Krippendorff's (2018) classical definition of content analysis as “a research technique for making replicable and valid inferences from texts and the contexts of their use,” adapted to audiovisual environments in which text, image, and sound coexist within a single narrative plane (Rose, 2016).

The proposed model is conceived as a staged methodological dispositif, designed to identify, measure, and correlate signs of the glorification of crime in Chilean urban music. More specifically, the procedure integrates three levels:

- (1) Lexical level: detection and quantification of meaning units (words and phrases) associated with weapons, drugs, crime, and luxury.
- (2) Visual level: recording and categorization of objects, settings, attitudes, and status emblems present in the music videos.
- (3) Relational level: analysis of collaboration networks among artists, understood as a plane of symbolic circulation in which shared imaginaries are disseminated and reinforced.

The multimodal approach responds to the principle of code interdependence (Kress & Van Leeuwen, 2001), according to which verbal and visual dimensions should not be examined in isolation, but rather as components of a single structure of meaning. In this article, the corpus of Chilean urban music functions as a pilot application, the purpose

of which is to verify the clarity of the operational definitions, the relevance of the categories, and the reliability of the procedure.

2.3. Epistemological approach

The study assumes that cultural products are not mere social reflections, but active forms of meaning production (Hall, 1997). From this perspective, urban music is not analyzed as a documentary record of criminal reality, but rather as a social text—an aesthetic narrative that translates the inequalities, aspirations, and contradictions of contemporary urban life.

From this perspective, the model integrates empirical systematicity (frequencies, distributions, reliability) with interpretative sensitivity (semiotic reading, contextual analysis), following Jensen's (2013) guidelines on mixed methods in communication research.

2.4. Corpus and selection criteria

The corpus consists of 12 music videos and 12 lyrics from Chilean urban music released between 2019 and 2024. Selection was conducted through purposive sampling aimed at ensuring diversity in styles, collaborations, and visual repertoires.

This corpus functions exclusively as a pilot application of the model. Its purpose is not to produce substantive conclusions about urban music, but rather to assess the operability of the lexical dictionary, the visual matrix, and the relational dimension. Selection decisions therefore respond to methodological criteria (semantic and visual diversity) rather than to statistical representativeness.

2.5. Research instruments

2.5.1. Lexical dictionary of semantic categories

The lexical dictionary structures the semantic fields relevant to the phenomenon under study through manual lemmatization and contextual review. It includes base lemmas, local variants (Corporación de la Lengua Española, 2023), anglicisms, borrowings from *lunfardo*, and compound expressions (Rincón, 2009; Sandoval, 2020).

Organization by semantic fields (crime, weapons, drugs, luxury/status) enables replicable lexical counting and subsequent comparison with visual patterns. The dictionary was subjected to a preliminary test with two coders in order to refine ambiguous or polysemic lemmas. Lexical counting was conducted through manual lemmatization and automated validation using scripts in RStudio (version 2024.09) to identify absolute and relative frequencies. This stage produced an occurrence table that made it possible to detect dominant semantic fields and their distribution across artists.

2.5.2. Visual coding matrix

The visual matrix functions as a codebook that records the presence, absence, and degree of predominance (high, medium, low) of objects, settings, attitudes, and gestures associated with repertoires of status, risk, and violence.

The categories include: luxury clothing and accessories; money; vehicles; weapons; substances; corporeality; and settings associated with neighborhoods or consumption. Decision rules were incorporated for borderline cases (ambiguous props, symbolism, irony, fictionalization).

The matrix was developed following the guidelines proposed by Rose (2016) and Ball and Gilligan (2010) for the analysis of images in digital media, ensuring consistency between visual units (frame-by-frame) and narrative dimensions.

2.5.3. Coding template

A template was designed that integrates the lexical register, the visual register, and coding decisions. This template allows each audiovisual unit (video and lyrics) to be systematized and facilitates the subsequent calculation of lexical-visual convergence and inter-coder reliability.

2.6. Analytical procedure

The analytical procedure was developed in four phases:

- (1) Data extraction and cleaning: collection of lyrics (Genius) and music videos (YouTube) in textual and audiovisual formats; orthographic normalization and segmentation by verses/scenes.
- (2) Lexical coding: manual and automated counting of terms; grouping by lemma and semantic category according to the constructed dictionary.
- (3) Visual coding: systematic observation of scenes based on the codebook; recording of visual predominance (1 = low; 2 = medium; 3 = high) for each defined category.
- (4) Triangulation: comparison between lexical and visual frequencies; calculation of convergence using a simple lexical-visual correlation index, and qualitative analysis of divergent cases to refine decision rules and categories.

This component constitutes one of the central contributions of the model, as it offers a replicable procedure for future research.

Information on collaborations between artists was obtained through the public Spotify API, using authenticated calls to the Spotify Web API (artist, track, and album endpoints) to collect metadata on songs, featuring practices, and links between performers. The data were processed in RStudio (version 2024.09), using the *httr*, *jsonlite*, and *igraph* libraries for data cleaning, normalization, and network visualization.

This stage made it possible to construct an artist-by-artist matrix and to generate the collaboration graph used in the relational analysis.

3. Validation of the lexical–visual model: pilot application

3.1. Overview of lexical frequencies

The lexical analysis of the 12 songs yielded a total of 4,238 lexical units relevant to the dictionary. Of these, 38.4% corresponded to the “luxury” field, followed by “weapons” (24.6%), “drugs” (21.7%), and “crime” (15.3%). This distribution indicates that the lexical dictionary is capable of consistently discriminating among differentiated semantic fields and of concentrating the coding on categories theoretically defined as relevant.

Table 1
Lexical frequencies by category

Category	Absolute Frequency	Percentage (%)	Standard Deviation
Luxury	1,627	38.4	12.3
Weapons	1,042	24.6	9.7
Drugs	920	21.7	8.9
Crime	649	15.3	6.8

From a methodological standpoint, these results confirm the operability of the instrument in capturing the recurrence of specific semantic repertoires in audiovisual corpora and provide a solid basis for subsequent comparison with the visual dimension of the model.

3.2. Visual distribution of categories

Observation of the music videos revealed a visual pattern consistent with the design of the codebook. The category clothing/status was present in 100% of the works, followed by money (83%), vehicles (75%), weapons (58%), drugs (50%), settings (42%), and attitude (38%).

Table 2
Visual presence by category

Visual Category	Frequency of Presence	High Predominance (%)	Lexical–Visual Concordance (%)
Clothing / status	12	91.6	84.3
Money	10	75.0	89.2
Vehicles	9	66.6	81.1
Weapons	7	58.3	76.4
Drugs	6	50.0	73.0
Settings	5	41.6	69.1
Attitude	4	33.3	63.7

From the perspective of the model, this distribution indicates that the visual matrix enables the systematic recording of both the presence and predominance of the defined categories, distinguishing between elements that are central to the staging (clothing, money, vehicles) and others with more limited occurrence. This confirms the applicability of the instrument to music videos characterized by a high density of signs, as well as its potential for comparative studies across other audiovisual contexts.

3.3. Lexical–visual convergence

To estimate lexical–visual convergence (LVC), an indicator was defined to measure the proportional correspondence between the relative frequency of a category in the lyrics and its operational presence in the music video. In the pilot application, the categories exhibited differentiated LVC values, with higher levels observed in “luxury” and “weapons,” and lower levels in “crime” and “drugs.”

From a methodological standpoint, these results show that the LVC indicator is sensitive to variation across categories and makes it possible to detect patterns of alignment and misalignment between verbal and visual dimensions. The experience suggests that this type of measure can be incorporated into future multimodal studies as a complement to qualitative analysis, offering a standardized way to compare the relationship between lyrics and image across different corpora.

3.4. Intercoder reliability

Intercoder reliability was assessed using Cohen’s Kappa coefficient, which is appropriate for nominal categorical variables. Unlike simple percentage agreement, Kappa corrects for agreement attributable to chance, providing a more rigorous measure of consistency between coders.

The value obtained for visual coding was $\kappa = 0.776$, which, according to the classification proposed by Landis and Koch (1977), corresponds to a “substantial” level of agreement. This result indicates that the operational definitions are sufficiently clear to produce convergent decisions among independent analysts.

Table 3
Intercoder reliability by category

Category	κ	Interpretation	Adjudication Notes
Weapons	0.86	Very good	High visual consistency in explicit scenes.
Drugs	0.74	Substantial	Difficulties in distinguishing “decorative” from “narrative” elements.
Crime	0.71	Substantial	Requires contextual anchoring (lyric verse/action).
Luxury	0.83	Very good	High agreement regarding visible objects and brands.
Settings	0.69	Acceptable	Adjustment of criteria for defining the “narrative environment.”

The values obtained confirm the robustness of the instrument and the clarity of the

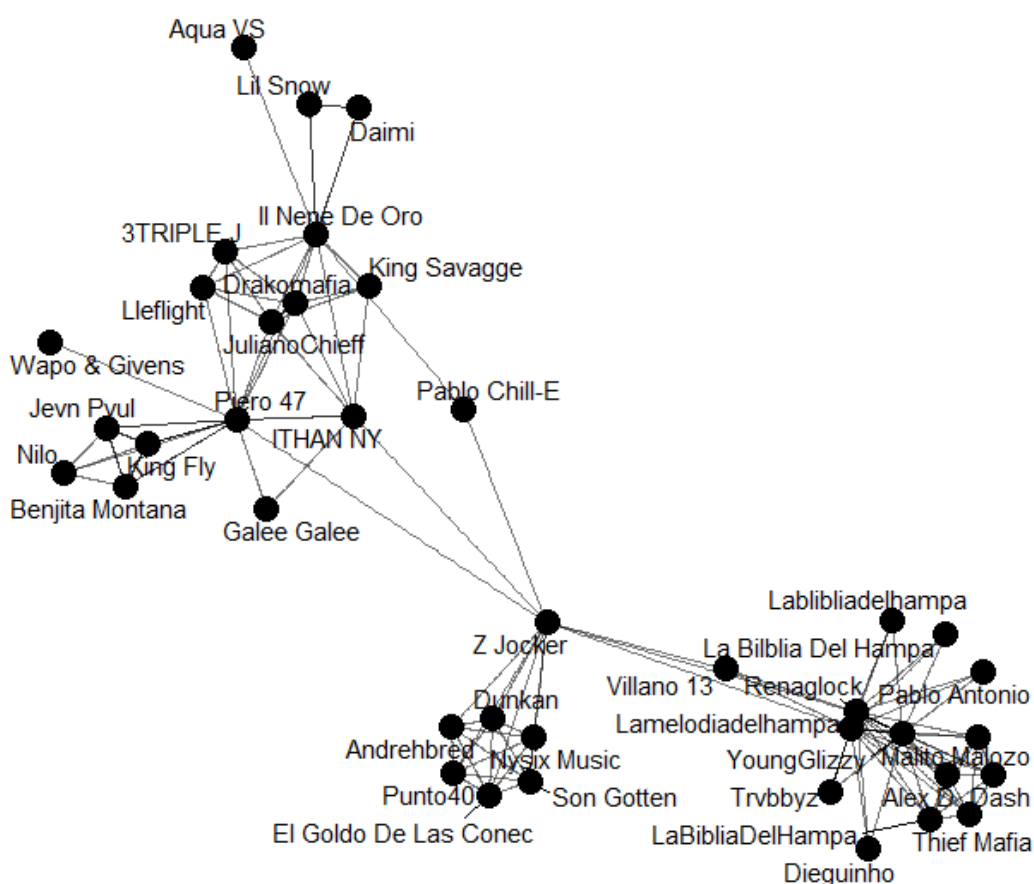
operational definitions. The most relevant differences were concentrated in the categories of settings and attitude, where coders tended to diverge in their interpretation of the intentionality of postures or environments (e.g., “neighborhood,” “party,” “warehouse,” “street”). Following a joint review, complementary decision rules were incorporated (e.g., “central presence,” “supporting element,” “ambient background”).

The validity of the model is further strengthened by: (1) lexical–visual convergence, which makes it possible to verify correspondences between semantic fields and visual repertoires; and (2) the internal consistency of the lexical dictionary, assessed through preliminary tests of lemmatization and variant selection.

3.5. Relational analysis: collaboration networks

Network analysis was employed in the pilot application to assess the relational dimension of the model. Based on credited collaborations among artists, a graph was constructed in which each node represents a performer and each edge represents a collaboration recorded during the study period.

Figure 1
Collaboration network graph



Network visualization made it possible to identify differences in node centrality and to overlay semantic information derived from the lexical dictionary (for example,

dominant categories associated with each artist), illustrating the potential of the relational component for studying the circulation of symbolic repertoires in audiovisual scenes. Rather than offering substantive conclusions about Chilean urban music, this stage demonstrates that the model is compatible with network analysis techniques and can be extended toward approaches that combine relational structure with multimodal patterns.

4. Discussion and Conclusions

4.1. Scope of the lexical–visual model

The pilot application of the lexical–visual model makes it possible to reposition the debate on phenomena such as the so-called glorification of crime onto a primarily methodological terrain. Rather than offering a closed interpretation of the analyzed corpus, the study demonstrates that it is feasible to translate complex discussions—crime, narco-aesthetics, inequality, status—into observable dimensions that can be systematically coded and compared. By simultaneously engaging with semantic fields and visual repertoires, the model provides an operational framework for analyzing how languages of transgression are articulated in contemporary audiovisual products.

The results of the pilot application show that the lexical dictionary and the visual matrix can capture regularities and nuances in the representation of transgression, luxury, and risk. The differential frequency of semantic fields and the distribution of visual categories are not presented here as “proof” of a substantive thesis about Chilean urban music, but rather as evidence that the model can detect patterns of recurrence, thematic hierarchies, and differences in the degree of explicitness between verbal and visual dimensions. In this sense, the primary contribution lies not in what the corpus “says” about the glorification of crime, but in how the model enables these repertoires to be observed, measured, and compared.

4.2. Convergences, dissonances and multimodal reading

The introduction of a lexical–visual convergence (LVC) indicator constitutes one of the model’s specific contributions. The possibility of estimating, in a standardized manner, the degree of alignment between what is enunciated in the lyrics and what is represented visually opens up a fertile space for multimodal analysis. Differences in LVC across categories show that the model is sensitive to the coexistence of convergences (when the image reinforces the dominant semantic field) and dissonances (when verbal and visual dimensions diverge or are displaced).

This behavior is not interpreted here as confirmation of a hypothesis about the “essence” of the urban genre, but rather as a demonstration of the model’s potential to map complex discursive operations: intensification, stylization, attenuation, or displacement of specific contents. In this sense, the proposal dialogues with the multimodal tradition that conceives the music video as a syncretic text (Kress & Van Leeuwen, 2001; Lotman, 1996), while contributing a concrete procedure for both quantifying and qualifying the relationship between different modes of signification.

The dual lexical–visual reading also makes it possible to articulate more precisely the turn proposed by authors such as Hall (1997) and Martín-Barbero (2003): representation does not merely “reflect” reality, but actively constructs it through codes, genres, and styles. The lexical–visual model offers a way to operationalize this premise by identifying how signs are organized into observable patterns and how these patterns can be compared across scenes, genres, or different time periods.

4.3. *Articulation with previous methodological literature*

The proposed model is situated within the tradition of content analysis (Krippendorff, 2018; Neuendorf, 2017) and visual methodologies (Rose, 2016; Ball & Gilligan, 2010), while introducing two shifts that are particularly relevant for the methodological field. First, it explicitly integrates the multimodal dimension, avoiding the treatment of text and image as separate objects of analysis. The articulation between the lexical dictionary and the visual matrix responds to calls by Jensen (2013) and other scholars to develop mixed methodologies capable of reconciling measurement and interpretation in communication studies.

Second, the incorporation of a relational dimension through network analysis extends the traditional scope of content analysis. By working with nodes (artists or other actors), edges (collaborations), and structural indicators (centrality, communities, density), the model makes it possible to link semantic and visual repertoire with patterns of symbolic circulation. Rather than offering an analysis of the music industry, this dimension demonstrates that tools from network theory can be integrated with multimodal analysis, generating a space for dialogue between approaches that are usually applied separately.

In this way, the model does not seek to replace existing methodologies, but rather to articulate elements from several of them into a coherent dispositif: the rigor of content analysis, the interpretative sensitivity of cultural semiotics, and the structural perspective of network analysis.

4.4. *Limitations of the model*

The pilot application also reveals a set of limitations that must be taken into account. First, visual coding continues to depend on the situated interpretation of coders. Although Cohen’s Kappa coefficient indicates a substantial level of agreement, the categories of *settings* and *attitude* presented greater difficulties in achieving consensus, suggesting the need for further refinement of operational definitions, the inclusion of additional examples, and, potentially, the development of more extensive training protocols.

Second, the lexical dictionary, even though it was subjected to preliminary testing, was constructed from a specific corpus and within a particular linguistic context. Its transfer to other cultural or linguistic settings requires adjustments, a review of local idioms, and likely the incorporation of new categories. The model is replicable, but not universal: it demands careful adaptations according to the characteristics of each scene.

Third, the relational dimension depends on the availability and quality of metadata on digital platforms. Changes in recommendation systems, tagging practices, or visibility mechanisms may affect the ability to reconstruct collaboration networks with the same level of detail. Moreover, integrating relational information with the multimodal dimension requires additional decisions regarding which ties are relevant for each study. Finally, validation was conducted on a limited corpus that functions as a proof of concept. Future research should apply the model to larger and more diverse samples and contrast its results with other methodological approaches in order to assess convergences, discrepancies, and blind spots.

4.5. Potential applications and future research directions

Despite these limitations, the lexical–visual model shows strong potential for application across multiple areas of communication research. Its structure allows it to be adapted to the analysis of music videos from other genres, audiovisual advertising campaigns, serialized fiction, television news on security or migration, and native digital platform content (reels, shorts, TikTok). In all these cases, the combination of a lexical dictionary, a visual matrix, and a relational dimension provides a set of tools for examining how imaginaries of risk, inequality, consumption, identity, or citizenship circulate and are reconfigured.

From a technical standpoint, a promising line of development lies in exploring the integration of the model with automated image and video analysis tools (computer vision), as well as with advanced natural language processing techniques. Such integrations could streamline lexical counting and visual coding while preserving a qualitative interpretative core capable of supervising and contextualizing the results.

In addition, the model could be linked to audience reception and participation studies by incorporating engagement metrics or social media interactions. This would make it possible to relate detected multimodal patterns to the ways in which audiences interpret, discuss, or reappropriate audiovisual content, thus extending the analytical focus from textual production to processes of circulation and resignification.

In sum, the validation conducted demonstrates that the lexical–visual model constitutes a robust and flexible framework for the analysis of audiovisual narratives. Its main contribution does not lie in definitively resolving debates on the glorification of crime, narco-aesthetics, or urban music, but rather in offering a set of methodological tools that enable these and other phenomena to be addressed in a more systematic, transparent, and comparative manner. By combining lexical counting, visual coding, reliability assessment, and relational analysis, the proposal contributes to consolidating a methodological field capable of responding to the challenges posed by contemporary forms of multimodal representation.

5. References

- Ball, M., & Gilligan, C. (2010). *Visual methodologies for digital culture*. SAGE.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
<https://doi.org/10.1177/001316446002000104>
- Corporación de la Lengua Española. (2023). *Diccionario de chilenismos y expresiones coloquiales*. Santiago: CLE.
- Deflem, M. (2020). Popular culture and social control: The moral panic on music labeling. *American Journal of Criminal Justice*, 45(1), 2–24.
<https://doi.org/10.1007/s12103-019-09495-3>
- Díez, S. N. (2020). *Sobre rap, trap y calle: imágenes y fenómenos*. Kamchatka. Revista de análisis cultural, 16, 93-128. <https://doi.org/10.7203/KAM.16.16627>
- Hall, S. (1997). *Representation: Cultural representations and signifying practices*. SAGE.
- Hartley, J. (2010). Silly citizenship. *Critical Discourse Studies*, 7(4), 233–248.
- Hjarvard, S. (2015). *The mediatization of culture and society*. Routledge.
- Jensen, K. B. (2013). *A handbook of media and communication research: Qualitative and quantitative methodologies* (2nd ed.). Routledge.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Ley que Sustituye la Ley N.º 19.366, que Sanciona el Tráfico Ilícito de Estupefacientes y Sustancias Sicotrópicas y Deroga Otras Disposiciones Legales, N.º 20.000, Diario Oficial de la República de Chile, 16 de febrero de 2005.
- Ley sobre Seguridad del Estado, N.º 12.927, Diario Oficial de la República de Chile, 19 de junio de 1958.
- Lotman, Y. (1996). *La semiosfera I: Semiótica de la cultura y del texto*. Cátedra.

- Martín-Barbero, J. (2003). *De los medios a las mediaciones: Comunicación, cultura y hegemonía* (3ª ed.). Anthropos.
- Mondaca, A. (2012). Narcocultura y poder simbólico: entre la violencia y la estética. *Revista de Estudios Culturales Latinoamericanos*, 9(1), 55–72.
- Neuendorf, K. (2017). *The content analysis guidebook* (2nd ed.). SAGE.
- Olivares, P. (2023). Estrategias de visibilidad en la escena urbana chilena: Plataformas, colaboraciones y capital simbólico digital. *Revista Comunicación Contemporánea*, 8(1), 22–39.
- Reguillo, R. (2000). *Emergencia de culturas juveniles: Estrategias del desencanto*. Norma.
- Rincón, O. (2009). *Los narcos también cuentan: Narcocultura y narrativas mediáticas*. Tercer Mundo Editores.
- Riquelme, S., Sule, V., Castillo, V., & Núñez, V. (2022). La periferia urbana a través de los lentes del trap chileno: narrativas juveniles. *Última Década*, 30(59), 140–172. <https://doi.org/10.4067/S0718-22362022000200140>
- Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials* (4th ed.). SAGE.
- Ruiz, O. (2022). The influence of narcoculture on popular music: A critical look at reggaeton's narco-messages and narco-representations. *Journal of the Royal Musical Association*, 147(1), 77–104. <https://doi.org/10.1017/rma.2022.13>
- Sandoval, M. (2020). La estética del exceso: Cultura visual y narcocapitalismo. *Cuadernos de Comunicación*, 5(2), 89–110.
- Vásquez, M. (2017). *Narcoestética y desigualdad urbana: Crónicas visuales de la periferia chilena*. Ediciones UDP.
- Wacquant, L. (2007). *Los condenados de la ciudad: Gueto, periferias y Estado*. Siglo XXI Editores.

Funding: no funding received.

Conflict of interest: none.

English translation: provided by the authors.

HOW TO CITE (APA 7ª)

Crisóstomo, P., Hernández, N., & Higuera, V. (2025). Lexico-visual Analysis and Lexical Frequencies in Urban Music: a Methodological Proposal for the Study of Audiovisual Narratives. *Comunicación y Métodos | Communication & Methods*, 7(2), 54-72. <https://doi.org/10.35951/v7i2.257>