

Publication status: This preprint has not been published elsewhere.

Distant Reading in the Arrabal: Women Cancionistas and Their Distinct Profiles in the Tango-Canción During Its Foundational Stage (1922-1935)

Aldo Mazzucchelli, Martina Bertone

<https://doi.org/10.1590/SciELOPreprints.15306>

Submitted on: 2026-03-05

Posted on: 2026-03-19 (version 1)

(YYYY-MM-DD)

Distant Reading in the *Arrabal*: Women *Cancionistas* and Their Distinct Profiles in the *Tango-Canción* During Its Foundational Stage (1922-1935)

Aldo Mazzucchelli (*) ORCID: <https://orcid.org/0009-0000-5895-0760>

Martina Bertone ORCID: <https://orcid.org/0009-0002-6415-823X>

(aldo.mazzucchelli@fhce.edu.uy, martina.bertone@fhce.edu.uy),

Universidad de la República (UdelaR), Facultad de Humanidades y Ciencias de la Educación (FHCE), Uruguay

Abstract This article analyzes the emergence of *cancionistas* in the foundational stage of the tango-canción (1922-1935) through Digital Humanities methodologies (Distant Reading). Countering the historical homogenization of the female voice in the genre, the research examines a representative corpus of the eight main female interpreters of the era, seeking to determine their enunciation strategies, their appropriation of linguistic registers (cultured vs. popular), and their thematic distribution according to the taxonomy proposed by Daniel Vidart. Using text mining algorithms¹ (TF-IDF with binary frequency), complex network modeling, and tri-polar gravity topologies, the study empirically demonstrates that the *cancionistas* did not form a uniform aesthetic block. The results reveal an active gender negotiation, polarized between the assumption of a male enunciation or "vocal transvestism" (41%) and the affirmation of a female lyrical voice, although the latter was generally also mediated by male lyricists (37.7%). Likewise, the lexical modeling exposes divergent discursive archetypes: from sentimental purism (Libertad Lamarque, Ada Falcón, Mercedes Simone), passing through satirical and *arrabalera* subversion (Tita Merello, Rosita Quiroga), to profiles with greater thematic breadth and varied lyrical voices (Azucena Maizani, Tania). It is concluded that the *cancionistas* operated as creators of their own repertoire, and by acting within the limits and horizons of expectation of their historical context, they actively began to rewrite the discursive boundaries of early tango.

Keywords: *Tango canción*, *Cancionistas*, Digital Humanities, Distant Reading, Gender, Lunfardo, Rosita Quiroga, Azucena Maizani, Libertad Lamarque, Mercedes Simone, Tania, Sofía Bozán, Ada Falcón, Tita Merello.

Declarations

Author Contributions: A.M.: Conceptualization, Methodology, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing. M.B.: Conceptualization, Investigation, Writing - Review & Editing.

Data Availability Statement: The textual corpus (*lyrics.corpus*), the classification metadata according to Vidart's taxonomy, the Python scripts used for topic modeling and the generation of network graphs, as well as the interactive versions (HTML) of the visualizations, are available for review and reproduction in the supplementary material attached to this article, at <https://doi.org/10.5281/zenodo.18881862>.

Conflict of Interest The authors declare that they have no conflicts of interest regarding the publication of this article.

(*) Corresponding author.

1. Introduction

The emergence of the *cancionistas* during the 1920s represents a fundamental moment in the configuration of the tango-canción (Dos Santos, 2011; Gobello, 2002; Sosa de Newton, 1999). Their irruption into an artistic and communicational universe controlled almost entirely by men—from lyricists and singers to theatrical and media impresarios—signified a disruption whose specific details of meaning it is imperative to investigate.

It is an accepted fact that, although tango lyrics may have existed during the *Guardia Vieja* (Old Guard) period (Assunção, 1984, pp. 188–190; Stilman, 1965, pp. 15–18; Tallón, 1964), the "*tango canción*" is born with "Mi noche triste" in 1917 (García Jiménez, 1964, 56). If that is the late beginning, by the final moment of the period we have chosen (1935), the tango is already the sung tango, and in song format—on records as well as on the theatrical stage and in cinema.

The eight "*cancionistas*" we will address in this work are far from having been the first female protagonists of tango. There are precursors whose role and importance have been well determined, such as Pepita Avellaneda, Linda Thelma, or the actress Manolita Poli (who precisely premiered the already cited first *tango canción*, "Mi noche triste"), among others. Néstor Pinsón (n.d.-a) reports and summarizes research by Hugo Lamas and Héctor Lucci confirming that the Frenchwoman Andrée Vivianne was the first woman who recorded tangos, starting on 2-minute cylinders between 1904 and 1908, and later moving to 25 cm records until the year 1909. Her registry completes nineteen recordings.² But the cited pieces, in a female lyrical voice, are an exception in a time dominated by bawdy ditties and lunfardo speech, adapted to an environment centered on music and dance. In general, if there were lyrics—which there almost never were—they were sexually bawdy, from a male point of view, direct, and often improvised. This is the most usual during the aforementioned era of the *Guardia Vieja*, which corresponds to the period of musical, orchestral, and dancing beginnings of tango in the Río de la Plata, along with its expansion to Paris, and which culminates between the beginnings of the *tango-canción*, with the consolidation of Carlos Gardel since 1917, and the formal revolution brought by Julio de Caro's orchestra since 1924 (García Jiménez, 1964, 71; Gobello, 2002: 41). The female singers who enter the scene from the 1920s occupy in the following period a role that will gradually diversify and expand.

The figures that will occupy us³ are Tania (Oct 13, 1895), Rosita Quiroga (Jan 13, 1896), Sofía Bozán (Nov 5, 1898), Azucena Maizani (Nov 17, 1902), Mercedes Simone (Apr 30, 1904), Tita Merello (Oct 11, 1904), Ada Falcón (Aug 17, 1905), and Libertad Lamarque (Nov 24, 1908). The vast majority—except for Mercedes Simone and Ada Falcón—entered the tango scene through the theater and singing on stage. To their gifts as singers, some of these artists added, to a greater or lesser extent, an attractive bodily presence that gave them the potential for exploitation in the various registers in which the image of the female body, increasingly less covered, gained centrality

on a massive level. During the period of about 14 years that we will address, the industry and creators took advantage of the direct impact these figures had in live theatrical performance to offer them the premiere of numerous pieces, on a massive and popular theatrical stage, dominated by the "*género chico*" (Barsky, 2021, p. 164; Mazziotti, 1982, p. 73), which would consolidate by those years in the popular "*teatro de revistas*" (revue theater). This, in addition to giving those involved the benefit of presenting novelties, served as a testing platform regarding the potential popularity of the new songs premiered in this way. The public's reaction, including the number of "encores" it demanded, were clear indicators of the convenience of investing in the recording of the new piece. But the *cancionistas* did not necessarily end up being the ones who actually recorded each of those pieces on disc; rather, the theatrical premiere and the actual recording responded to complex interests, commitments, and links within the industry and the milieu. By integrating themselves into this dynamic, the women were not merely passive interpreters, but active agents who navigated, experimented with, and shaped a lyrical ecosystem predominantly written by men. The *cancionistas* all integrated themselves, in their own right, into the phonographic recording market and contributed significantly to the massive diffusion of the tango canción, and—thanks to their appeal as a stage presence—they served as impact anchors for the premiere and testing of new hits—mostly written by men. In this giving voice to themes of a male constellation of lyricists—regardless of whether each lyric was written with a male or female lyrical voice—a negotiation was installed at the level of the imaginary and the industry where the female voice and presence modified the tango canción from its very beginning.

In this first stage of our research, the objective is twofold: on the one hand, to determine the presence of a female lyrical voice in the lyrics interpreted by women in that period; on the other, to trace in the *cancionistas* the appropriation and use of both the "cultured" modernist register of several important lyricists (men or women) (Vilariño, 1962; Mazzucchelli, 2006), and the "popular language" that already synthesized strands that have been defined and named in various ways: "*gauchesco*" and "*lunfardo*" (Vilariño, 1962, p. 17), "*orillero*" (Rossi, 1958, pp. 277-285), "*lunfardesco*" (Vidart, 1967, p. 64), etc. These are the strands present in the initial codification of the tango lyric, as we will discuss. Likewise, we review in a general way the thematic axes addressed by these *cancionistas*, taking as a reference the classification by Vidart himself (1967). The methodological foundation for the choice of these variables is discussed in the corresponding section.

To proceed with this initial stage, we have delimited a chronological period and selected the eight main *cancionistas* of that era—already mentioned—extracting a representative corpus of between six and eight songs per interpreter. This volume, in our opinion, is broad enough to yield preliminary conclusions. Although we declare the preparatory character of this study, the results already reveal the use of a mixed palette in terms of lyrical voice, the audacious assumption of popular language, along with the integration of the cultured and modernist stylistic heritage, by key figures, and a marked distinction of profiles ("archetypes" of the *cancionista*) based on their artistic and image choices. All these factors contributed not only to installing the foundational presence of women in tango but also to consolidating repertoires and themes that sometimes reproduced male stereotypes, and at other times gradually built a female voice of their own. In this sense, the decision of each interpreter arises from her own convictions and choices within a conditioned

environment. A specific zone of the corpus shapes the incipient existence of what could be called a "female tango". We affirm this after several conceptual caveats: overwhelmingly male writing; a distribution of imaginaries in which, for the woman, comicality operates perhaps as a complement to the figure of the "*guapo*" and male violence, without questioning it; ⁴ the weight and emotional nuances of the interpretation that transcend the lyrics and "feminize" themes also sung by men; the physical and sexualized image of the woman on stage and in cinema; and incursions by some of the *cancionistas* into challenging zones for the established female image. Referring to the same artists that occupy us, Estela Dos Santos criticizes them by observing that the "*great creators of style... of the 1920s... were able... to adopt gestures of command and point directions. And they did not know how to do so.*" (2011, p. 11). However, from our corpus it emerges that these creators differentiated themselves sharply from one another and successfully explored varied routes, which suggests already a restlessness, far removed from that supposed taming and acceptance of a solely male ideology. Perhaps, then, one must avoid once again any risk of an evaluative anachronism that judges the femininity of women in the 1920s by transferring sociological and ideological criteria corresponding to the present, and instead listen to and explore the routes that they themselves effectively managed to open for their time and circumstance.

2. Methodology

In a first analysis, we must therefore distinguish at least three fundamental dimensions of the phenomenon:

a) *The lyrical voice*: The grammatical definer of who utters the text. b) *The content and point of view*: A semantic dimension that is not automatically attributed to the gender of the author or the interpreter (for example, a tango written by a man in a female grammatical voice may differ, or not, from one written by a woman on the same subject). c) *The semiotic presentation of the interpreter*: This depends decisively on the broadcasting medium employed (record, theater, cinema, or radio), and in that sense must be unfolded into two levels. On the one hand, (c1) *the auditorily perceived gender*, that is, the sound register and timbre (male or female) of the vocal interpretation; and on the other, (c2) *the visual and performative dimension* (the set of signs in a live or on-screen presentation). This last scenic level is also neither simple nor univocal (for example, the reception of a woman dressed as a gaucho is radically different from one attired in a party dress).

Added to the above, there is a decisive factor that has nothing to do with the lyrics, but with the sensitivity communicated in its interpretation (for example, Sofía Bozán's interpretation of "Yira, yira", protective and compassionate, is very different from Carlos Gardel's, sarcastic and hopeless, even though both sing exactly the same lyrics). This can only be perceived and commented on in the experience of the music, and is impossible to systematize with tools of commentary or digital hermeneutics.

These initial difficulties cannot all be addressed at once. It is necessary to delimit the object of analysis to one or some of the aforementioned factors, in order to generate an initial field of clarity, upon which to build future levels of *distant reading* (Moretti, 2005). The lyrics of each song have been incorporated after verifying, through direct listening, the exact lexical and grammatical details

employed by the *cancionista*, guaranteeing an objective and stable factor. Regarding the gender of the singer, both the specific style of each singer and—more generally—the almost absolute use of the light soprano register—with the exception of Mercedes Simone, who often ventured into the mezzo-soprano register—markedly differentiated these recordings from the male versions, dominated then by the tenor register (Agustín Magaldi, Ignacio Corsini, and the early Gardel, before his shift toward a more resolutely baritone voice in the final part of his career).

2.1 Construction and Foundation of the Corpus

To build our initial corpus, we had to make two decisions: who should be included and what texts to analyze from each one.

Regarding the first, there is sufficient evidence and consensus that the eight selected interpreters are the key *cancionistas* of the period.⁵ These eight women were the ones with the greatest popularity, the ones who occupied the priority spaces in theater, radio, record, and cinema, and the ones who gave the female register its initial place in tango.

Regarding the second—the selection of the songs—there are two considerations. The size of the sample per *cancionista* is the first and fundamental one. This is currently small—we plan to expand it in a longer-term study—and it attempts to concentrate by carefully choosing the most characteristic songs to determine, in this primary study, simply the typical profile of each of them for the defined variables. The second consideration, which arises from the first, is how to substantiate the choice of the selected songs. Such justification cannot be based on official sales figures due to the institutional precariousness of the time. To determine the dates and matrices we resorted to the DAHR (Discography of American Historical Recordings) database, essential for US labels like Victor, but which almost entirely omits European and local labels, such as the vital Nacional-Odeón (managed by Max Glücksmann).

To the lack of access to complete archives of competing labels is added the legal chaos of the River Plate musical market of the 1920s. Authorial and royalty information was only unified and concentrated with the creation of SADAIC on June 9, 1936. Although prior entities existed such as the Círculo de Autores y Compositores de Música (since 1920) and the Asociación Argentina de Autores y Compositores de Música (since 1930), which sought to provide protection for creators' rights, the Intellectual Property Law No. 11.723 was only sanctioned on September 26, 1933, and none of the previously operational societies had the real power to order a lucrative but chaotic market, plagued by sheet music forgeries, pirate reproductions, and imprecise attributions.

Therefore, given the material impossibility of rigorously ranking the songs by sales volume, the inclusion of these specific songs is based on an argument of shared authority. The selected corpus reflects a certain historiographical consensus on which were some of the most popular, representative, or foundational pieces of each interpreter. This selection is transversally supported by the most accepted research on the subject (such as the works of Estela Dos Santos (2021), the encyclopedic knowledge of José Gobello (2002), documentary reference bases like the magazine *El alma que canta* (1916-1961) by Vincenzo Buccheri, which published lyrics and sheet music giving

a certain indication of initial impact, plus the biographical sketches and comments routinely included in different general histories of tango, and the reviews and mentions gathered in recognized established sites like the *Todotango* site and similar ones. It is slippery ground, and in no way do we affirm that our selection is unobjectionable or complete. But it constitutes, as far as we know, the broadest sample of analysis of tango-canción sung by women in the initial era of the genre.⁶ It is enough for us to assure that the titles included for this first sample are representative of the popular imaginary of each *cancionista* in the studied stage. We believe that the findings found affirm and refine primary hypotheses about the image and role of each *cancionista* considered.

To avoid the possibility of overinterpreting statistically non-significant results—especially at the level of each isolated *cancionista*—we have tried to proceed to a reading that highlights the most notorious and general features, mentioning only in passing some nuances that will require a larger corpus to be confirmed or disproved.

A final caution to introduce has to do with the phenomenon of the factors of popular success of a song, which in no way can be attributed exclusively to its lyrics. Therefore, other elements have necessarily intervened for these songs, which enter our canon for each *cancionista*, to have triumphed. But we cannot, in this case, do anything other than concentrate on the lyrical factor as one of the ingredients of that success. Anyway, the lyrics of the *tango-canción* have not been a mere secondary element since they appear, but always cited, remembered, and central to the genre—including its most elaborate stage in the '40s, so our concentration on them is providing an important component of the cultural phenomenon.

As for the period, it deserves a separate discussion. Let's go to it.

2.2 Rationale for the period: 1922-1935

Our focus centers on the foundational stage of the popular consolidation of the genre, which we demarcate between 1922 and 1935. The starting point of the corpus is anchored in the first recordings in our selected group that—along with the theater⁷ and radio—massified the female voice in tango. For its part, the closing of this time window is marked by the convergence of three substantial transformations: a media one, a political one, and an individual one.

The *media change* is given by the massive irruption of sound cinema (a milestone marked by the premiere of the film *Tango!* in 1933).⁸ The cinema incorporated tango from its beginnings, granting it a different platform with a radically greater reach to the River Plate artistic world, whose image was until then confined to live performance and static photographs. The rise of sound cinema was meteoric starting from that beginning in 1933. The other media—especially radio and theater—had to adapt to this new media ecology. Female figures had to readjust and integrate their specific difference into the visual language, which generated a significant reorganization of the star system: some adapted very effectively (the paradigmatic case of Libertad Lamarque), others made incursions from varied and often contrasting angles, such as Azucena Maizani, Mercedes Simone, Tita Merello, or Sofía Bozán, and others did not manage or did not seek to adapt to the new

medium, such as Rosita Quiroga, or Ada Falcón (Wolff, 2003). The scene of the 1920s was, consequently, recoded.

The *political change* obeys a drastic institutional turn in the Argentine cultural landscape, driven by the systematic censorship of popular language. As Varela (2005, pp. 88-93) and Barsky (2016, pp. 55-79) point out, the public sphere began to be directly intervened by power in the early 1930s. Following the 1930 military coup, the de facto presidency of José Félix Uriburu initiated a moralizing crusade aimed at stopping the installation and diffusion of "popular language" in artistic expressions (ordering, for example, the alteration in 1931 of the title of Cobián and Cadícamo's tango "La casita de mis viejos" to the more formal "La casita de mis padres"). This cultural control culminated in August 1935—shortly after Gardel's death—with the publication of the "Instructions for Broadcasting Stations", a regulatory framework that explicitly prohibited "*idioms that bastardize the language*" and low-tone comicality based on harsh voices. The environment became increasingly restrictive, which means that the transition of the female interpreters to the big screen corresponded to a lyrical language that had already been heavily "domesticated".

Finally, the death of Carlos Gardel in June 1935 left an undeniable void. The consolidation of his artistic legacy opened a space both for imitators and for movements of formal innovation. The disappearance of his figure was as imposing as it was liberating, inaugurating a new stage. These three phenomena converge between 1933 and 1936—a period in which, furthermore, the authorial issue begins to be legally ordered with the creation of SADAIC—making it legitimate and necessary to establish the cut in that temporal environment. From there, tango mutates into a massive industry dominated by the large orchestras, exhibiting greater lyrical and musical elaboration, but subject to unprecedented state control over popular culture.

2. 3 Three analysis tools: the lyrical voice, the language, and the themes

For this study we have chosen to concentrate on three elements that allow us to characterize the profile of each of the *cancionistas*, and of the "female tango" as a whole, for the studied period.

2.3.1 - The lyrical voice

This is the simplest, apparently. It is a matter of first distinguishing, for each song, whether the emission comes, semantically, from a female, male (which we call "vocal transvestism"), or neutral lyrical voice—the latter without gender markers, something common in lyrics of a descriptive or philosophical consideration.

2.3.2 - The language, cultured or popular?

Whoever confronts the corpus of tango lyrics encounters a peculiar and heterogeneous linguistic object. By the 1920s, various layers had amalgamated into a repertoire that the people and artists used with a fine awareness of differentiation: "orillero" speech (which includes lunfardo and the speech of the urbanized peasant), insertions of immigrant languages (predominantly Italian), and borrowings—generally used in ironic ways—from French. Added to this is the contribution of the

lyricists, often heirs to the modernist or late-romantic poetry of the late 19th century, who, ignoring the River Plate avant-gardes, tried to adapt the underworld to cultured idioms.

To characterize the language of the texts in our corpus, we have taken as a reference the circumstantiated discussions on the different linguistic strands of tango present in Rossi (1958, p. 277 - 285), Vilariño (1962, p. 17-35), and Vidart (1967, pp. 59 -111), as well as Mazzucchelli's (2006) discussion on the drift and integration of romantic and modernist language with the tango canción. Rossi already points out, in his first edition of the year 1926, that the phenomenon of a underworld slang (for Buenos Aires, "*lunfardo*") exists in all cultures, and in none is it considered the "national language," and instead highlights the phenomenon of synthesis of many strands that is forming the popular River Plate speech in those years, and which flows into the lyrics of the *tango-canción*. Vilariño (1962, pp. 17-35) clearly distinguishes three lines that she calls "*lunfardo*", "*gauchesco*", and "*depurado*" (purified). Daniel Vidart (1955, 1967 p. 59-110), after a detailed linguistic and stylistic study, characterized the popular strand expressed in the tango lyrics of the period as an already mature synthesis for that initial moment of the tango canción, calling it simply "popular language" ("*lenguaje popular*"), clearly distinguishing it from strict *lunfardo*, and also from cultured language, to which he assigns a minor role in the whole. Stilman (1965, 37) calls it "popular speech" ("*habla popular*"), in a similar sense.

In our study we will therefore distinguish, in principle, two major types in the tango lyrics for the considered period. On the one hand, there is that strand of popular speech or language defined by Vidart, which synthesizes and includes originally *orillero* idioms, Italianisms, Gallicisms, and the maximum freedom to mix them in original ways. The imprint of a popular poet like Evaristo Carriego is often cited as a predecessor to this approach, which was brought to its culmination by masters of the genre like Pascual Contursi, Celedonio Flores, or Enrique Santos Discépolo. Flores is an exemplary figure due to his evolution, since from a *culterano* poet he went on to become a *lunfardesque* lyricist, after writing his "manifiesto", "*La musa mistonga*", premiered with resounding success by Rosita Quiroga. On the other hand, there existed a lyrical vein characterized by a markedly "literary" language, heir to modernism and romanticism in its late Hispanic version. Many of the main or most prolific lyricists of the era had been trained—and were still under the literary, lexical habits, and emotional topics—of the poets of past decades, especially in the sequel of Rubén Darío or Leopoldo Lugones, or the neo-romantic impact of the very popular and eclectic Almafuerde. This second strand generates a type of lyricism quite clearly distinguishable from the first.

Faced with the braid of this complex lexicon and its frequent ironic or satirical use, a challenge arises: How to determine the belonging of a lyric to one or the other type? The methodological answer adopted in this work is the prior human curation of the corpus, both to determine the lyrical voice, and the type of language.

2.3.3 Semantic Codification Criteria: Daniel Vidart's Taxonomy

Thirdly—in addition to the gender of the voice, and the general typology of language employed—the question of the themes addressed arises. To provide the algorithm with context, we have applied to each text a rating based on the thematic matrix consolidated by Vidart himself.

To categorize the semantic universe of the 61 analyzed works, it was chosen to adopt and adapt the exhaustive taxonomy proposed by the Uruguayan anthropologist and essayist Daniel Vidart. This classification proves methodologically suitable due to its deep roots in the historical sociology of the Río de la Plata, allowing the mapping of the lyrics not only as literary artifacts, but as symptoms of social tensions.

The corpus was subjected to a process of close reading prior to its computational ingestion. Each composition was assigned a numerical code corresponding to Vidart's ten thematic macro-axes: (1) Peasant, (2) *Orillero*, (3) Urban, (4) Amorous/Sentimental, (5) Environment, (6) Satirical, (7) Playful, (8) Philosophical, (9) Social, and (10) *Metatango*.⁹

Nevertheless, tango lyrics tend, by their own poetic nature, toward hybridization and the overflowing of categorical boundaries. An amorous complaint (Theme 4) can, for example, be inextricably anchored in the description of a cabaret (Theme 5) or be tinged with a strong metaphysical skepticism (Theme 8). To mitigate the reductionism inherent to any labeling system, the codification was designed with a double-entry architecture.

On the one hand, a Main Category (*Cat*) was defined, which captures the dominant plot vector of the work. On the other hand, the registration of a Subcategory (*Subcat*) was enabled, which operated in two distinct but complementary methodological ways:

- **Thematic Intersection:** When the lyric pivoted equitably between two macro-axes (e.g., a simultaneously *orillero* and satirical tango), a second number was recorded, allowing the work to statistically feed both orbits in the network models.
- **Internal Granularity (Subtypes):** When the nuance required specificity within the same macro-theme, letters (A, B, C, D, E) derived from Vidart's own analytical breakdown were used. This allowed, for example, that within the hegemony of the Amorous Theme, the algorithm could distinguish in the future between the narrative of abandonment, the nostalgia for lost love, or betrayal; or else, within the Satirical Theme, separate the mockery of the social upstart from the irony regarding fashions.

In sum, this decision of dual codification (hermeneutic and structured) guaranteed that the subsequent data mining operated on a metadata base rich in nuances, capable of respecting the ambiguities of the poetic text and avoiding the watertight compartmentalization of the *cancionistas*' repertoire. This prior manual classification allows the subsequent text mining to cross the lexical and grammatical frequencies with a solid thematic pre-definition, thus bypassing the semantic blindness of purely computational analysis.

2.4. Applied Computational Methodology

2.4.1. Computational Design and Analytical Visualization (Distant Reading)

To transcend isolated hermeneutic reading and capture the systemic dynamics of the corpus, this work implements a macroanalytical methodological design (Jockers, 2013) based on text mining and complex network analysis. The computational processing was entirely developed in the Python programming language, employing updated data science libraries (`pandas`, `scikit-learn`, `networkx`, and `plotly`).

The computational processing was executed through Python scripts. The source code was developed and calibrated for this project with the assistance of artificial intelligence language models, under the philological direction of the authors. We selected and calibrated the algorithms to respond to specific philological problems based on the objectives of the work, this being finally structured into four phases of visual modeling:

A. Systemic Flow Modeling (Sankey Diagram) To map the ecology of the tango interpreted by women, we required a topology that showed the intersection of simultaneous categorical variables. An interactive Sankey Diagram (or alluvial graph) was chosen. This visualization allows tracing the "flow" of the repertoire from the origin node (the *cancionista*), passing through her enunciation decisions (Lyrical Voice and use of Lunfardo), until emptying into the thematic sinks (Vidart's categories). The thickness of the links reflects the statistical weight, allowing identifying at a single glance ordered aesthetic profiles versus more hybrid or chaotic repertoire strategies. The product is included at the end of our analysis, by way of a synthetic gathering of the findings after the analytical phase of the work.

B. Lexical Extraction and the Refrain Problem (TF-IDF with Binary Frequency) Traditional word counting (raw frequency) is ineffective in popular lyrics due to the "refrain effect", where the mechanical repetition of a phrase distorts the semantic relevance of the term. To solve this, we applied the TF-IDF (Term Frequency – Inverse Document Frequency) mathematical model.

A crucial methodological decision was made: to apply a binary frequency to the algorithm. In this way, the machine ignores whether a word appears one or ten times within the same file; it only registers its presence (1) or absence (0). The TF-IDF model severely penalizes transversal lexicon (words common to all tangos) and exponentially rewards unigrams and bigrams that are exclusive and distinctive of a thematic category or of a particular *cancionista*.

C. Lexico-Thematic Archetypes (Force-Directed Network Graphs) To empirically show the archetype of each *cancionista*, semantic networks were built using graphs directed by physical algorithms (Force-Directed Graphs through the `networkx` library). Instead of tabular lists, the graph places the interpreter as a massive central node. Orbiting around her are the "thematic nodes", whose size is directly proportional to their frequency in the artist's repertoire. In turn, connected to these themes, gravitate the lexical satellites (the distinctive words or syntagms extracted via TF-IDF). The repulsion algorithm (Spring Layout) organizes the space, revealing visual constellations

that demonstrate how dissimilar interpreters construct radically different semantic universes, validating the hypothesis of a conscious divergence in the construction of the tango "self".

D. Topology of Gender Negotiation (Tri-Polar Systems) Finally, to address the phenomenon of vocal transvestism and gender negotiation, statistical linearity was abandoned in favor of a Tri-Polar Gravity model. Three massive nodes were anchored in the two-dimensional plane, representing the immovable poles of enunciation: Male, Female, and Neutral Voice.

By introducing the nodes of the *cancionistas* and the themes in this space, the algorithm calculated the "gravitational pull" based on the number of songs associated with each voice. The result is a topological map where interpreters and themes self-organize spatially. Those women who create or adopt a female-emitter text, or adapt the original text toward a female emitter, are drawn toward said pole, while those who operate under strict mimesis of a male emitter remain suspended in the orbit of the male voice. This map visually demonstrates that the lyrical gender in early tango was not a given condition, but a force field in constant tension and negotiation.

4. Results

4.1. Thematic Ecosystem and Gender Negotiation: A Quantitative Approach

To transcend the isolated reading of the compositions and capture the systemic dynamics of female tango lyrics in its foundational stage, this first section presents a statistical and topological analysis of the total corpus (n=61). Through data visualization techniques (Distant Reading), three fundamental descriptive dimensions are exposed. Firstly, the general thematic hierarchy is modeled through a Sunburst (multilevel) chart, which allows quantifying the hegemony of Vidart's macro-axes without losing the granularity of their subthemes. Secondly, a Donut chart exposes the percentage distribution of the lyrical voice (male, female, and neutral), evidencing the slight superior quantitative weight of "vocal transvestism", while at the same time a significant female identity emerges as an enunciation strategy.

4.1.1 Thematic Distribution of the Corpus: Sentimental Hegemony and Satirical Counterweight

The hierarchical visualization of the total corpus reveals a strongly asymmetrical thematic topology, which suggests a certain delimitation of the discursive territories assigned to—or appropriated by—the *cancionistas*. To a large extent, the lyrical ecosystem is dominated by Theme 4: Amorous, which concentrates almost half of the analyzed works (46%, equivalent to 28 songs). This statistical preeminence would seem to indicate that, in its foundational stage, the phonographic industry and the public's horizon of expectations tended to channel the female voice preferably toward the framework of intimacy, affective conflict, and the sentimental narrative. The internal fragmentation of this macro-group (where the Subtheme of the Betrayal of Love with 15% and the Subtheme of the Sorrow of Love with 10% stand out) underlines, however, that this sentimental space did not operate in a monolithic manner, but housed a certain typology of nuances, profiles, and romantic outcomes.



Figure 1. Hierarchical Distribution of the Corpus (Sunburst Model). Multilevel representation of the 61 tangos analyzed according to Daniel Vidart's taxonomy. The inner ring shows the hegemony of the macro-axes (predominance of Theme 4: Amorous and Theme 6: Satirical), while the outer ring breaks down the granularity of the specific subthemes addressed by the *cancionistas*. Note: Key terms mapping for the original graph: Amorous = Amorous; Satirical = Satirical.

As the main counterweight to this sentimental hegemony, Theme 6: Satirical emerges, encompassing a not inconsiderable 23% of the corpus (14 songs). This solid presence demonstrates that the female repertoire was not limited to the passivity of the amorous complaint, but rather the interpreters systematically occupied a discursive space linked to mockery, irony, the critique of customs, the "cachada" of stereotypical characters (conveyed especially through its Subthemes "Critique of stereotypes" and "Critique of vital situations").

Finally, the periphery of the chart exposes a minoritarian but philologically significant thematic diversification. Categories like Theme 5: Environment (8%), Theme 8: Philosophical (8%), and Theme 10: Tango as a theme (Metatango) (8%) operate as exploratory veins.

Although the woman's voice tended to be circumscribed to the amorous or satirical axes, these peripheral presences suggest that the interpreters were not excluded from existential reflection, the painting of an era, or tango metaliterature.

4.1.2. The Lyrical Voice: "Vocal Transvestism" and the Dispute for Enunciation

The second descriptive dimension fundamental to understanding the matrix of this corpus is the distribution of the enunciative voice. The statistical analysis captured in the Donut chart reveals the existing tension between the adoption of the patriarchal matrix of tango and the installation of a female lyrical "self" by the *cancionistas*. Such mimetic hegemony is strongly disputed by the Female Voice, which reaches a highly significant 37.7%. The close percentage proximity between

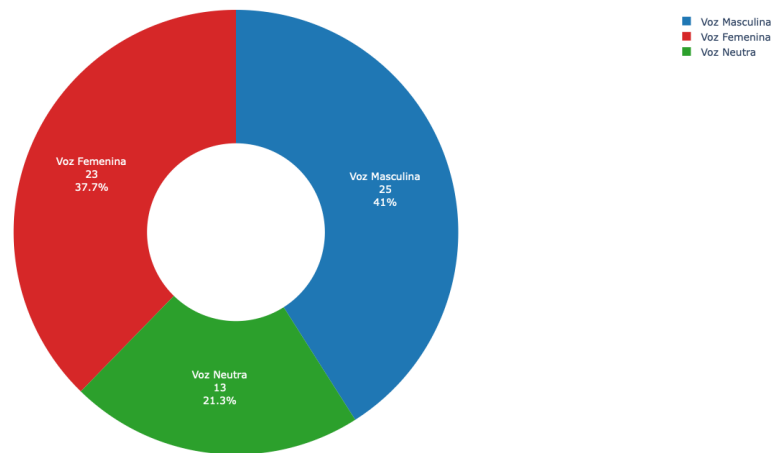


Figure 2. Distribution of the Lyrical Voice. Proportion of the corpus according to the grammar of enunciation. The tension between "vocal transvestism" (Male Voice, 41%), the affirmation of a female "I" (37.7%), and the emergence of ambiguity (Neutral Voice, 21.3%) is observed. *Note: Key terms mapping for the original graph: Voz Masculina = Male Voice; Voz Femenina = Female Voice; Voz Neutra = Neutral Voice.*

both poles shows the existence of a negotiation in the role of emitter, although in this time the male generic coordinate still continues to impose itself, while the portion of the corpus where the interpreter enunciates from an explicitly female "self", whether by ad hoc composition or by deliberate feminization of the text, is the field where the agency of these artists would potentially begin to consolidate. The data suggests that the *cancionistas* were not mere passive reproducers, but rather pulled almost 40% of their repertoire toward the construction of a female subjectivity of their own, within tango.

But this closeness between the percentages must not, certainly, be read uncritically as an installation of an "authentic and female" self, whatever that may be. We must keep in mind that the overwhelming majority of the considered lyrics (57 out of 61) are signed by men, and therefore even when the emitting self is grammatically female, one could speculate about the degrees and dynamics of generic assignment and appropriation based on a detailed semiotic analysis that would include sign considerations even beyond the lyrics themselves, something that remains for the moment outside the objective of this work.

The cold—but significant as a general framework—data yielded by this distribution is the relative preeminence of the Male Voice, which encompasses the largest portion of the corpus with 41%. This percentage shows with simplicity the predominant existence of what could be called "vocal transvestism" in the considered period. The presence of the woman occurs in negotiation with the voice of the man. Statistically, the majority strategy of the early interpreters consisted of executing the repertoire in strict mimesis with the original male perspective—the abandoned male, the *compadrito*, the *arrabalero* observer. This assumption of the male mask operated, to a large extent, as a validated aesthetic and commercial code that allowed women to enter a genre of markedly biased roots without altering its grammatical coordinates. We repeat: the illocutionary effects of the phenomenon are of interesting consideration, although—once again—outside our objectives here.

Finally, the chart exposes 21.3% of compositions executed in Neutral Voice. Although a minority compared to the two explicit gender poles, this space retains an interesting philological value. It groups those lyrics where the elision of the subject or the absence of grammatical markers enabled a sort of strategic ambiguity. In this terrain, the effect circumvented the imposition of "vocal transvestism" without a textual feminization taking place either.

In synthesis, the almost even polarization between male mimesis and female affirmation would seem to indicate that the grammatical gender within the lyrics in this period does not correspond to a unidirectional imposition, but to a space of symbolic negotiation.

4.2. Lexical Archetypes: The Semantic Imprint through Filtered Constellations

To demonstrate the radical divergence in the profiles of the eight *cancionistas*, we proceeded to extract the lexical matrix of their repertoires by means of network graphs. Methodologically, two extraction models based on the TF-IDF (Term Frequency – Inverse Document Frequency) algorithm were evaluated: one oriented to "bigrams" (two-word syntagms) and another to "filtered unigrams". For this definitive analysis, the unigram model applying a binary frequency parameter was chosen. This decision proved effective in resolving an inherent bias in popular lyrics such as the "refrain effect". While traditional counting (and bigrams) drag the statistical noise of words mechanically repeated in refrains, the filtered frequency registers the presence of a term a single time per song, regardless of its repetitions. By neutralizing the distortive weight of the refrain, the algorithm distilled a statistically distinctive vocabulary for each interpreter compared to the rest of the corpus, suggesting—although due to the size of the individual subsamples, the results at the lexical level cannot yet be considered significant—that behind the apparent uniformity of the tango of the twenties and thirties, each *cancionista* seems to have been constructing a semantic universe and an identity archetype in some singular way.

The most significant result of this preliminary study is, in our opinion, the determination of clearly differentiated spaces and thematic profiles for each of these *cancionistas*. Let's examine them.

4.2.1. Lexico-Thematic Constellations: Divergence and Singularity in the Archetype of the Cancionista

The visualization of the thematic repertoires through network graphs (Constellations of filtered unigrams) constitutes interesting empirical evidence against any notion of a homogenization of the woman's role in early tango.

Upon observing these architectures, it is first necessary to understand the physics of the algorithm that draws them (Force-Directed Graph): the distance of a thematic "planet" with respect to the central "Sun" (the interpreter) responds to a kind of statistical "law of gravity": massive and close themes possess a greater attractive force, because they concentrate the bulk of the repertoire and, therefore, a dense network of shared lexicon that "pulls" the node toward the center. Conversely, the smaller and more distant balloons in the orbit represent occasional incursions; they are semantic satellites of lesser density, which the algorithm pushes toward the periphery so as not to overlap them with the artist's gravitational core.

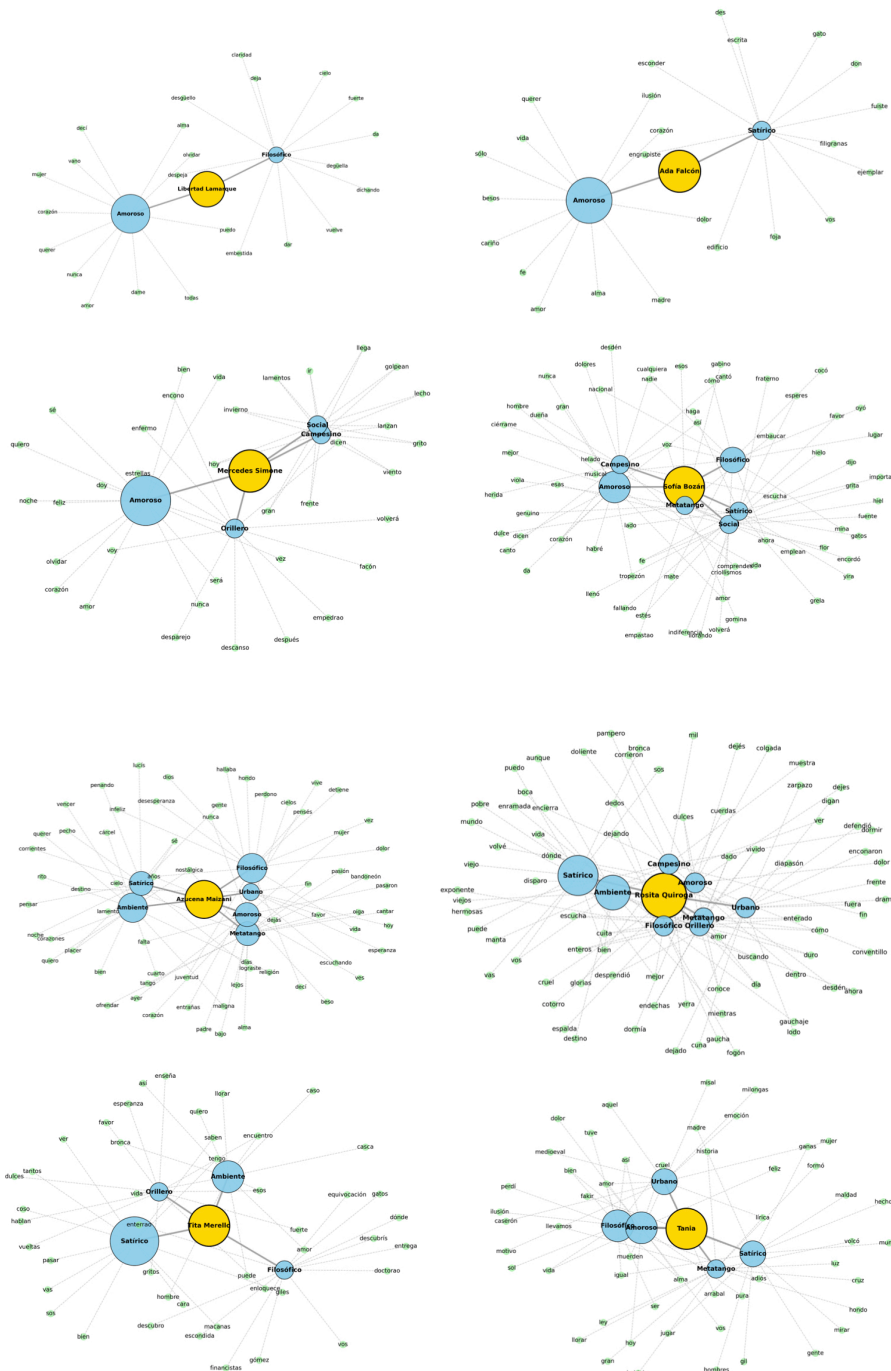


Figure 3. Comparative Matrix of Lexico-Thematic Constellations.

Simultaneous visualization of the eight analyzed profiles. Note the structural divergence between architectures of high concentration and low lexical density (e.g., Lamarque, Falcón) versus profiles of thematic dispersion and saturation of popular vocabulary (e.g., Quiroga, Maizani, Merello).

Under this topology, the exhibition of keywords (previously filtered of repetitions by refrains) operates as a symptom. What Vidart classifies as "Amorous Theme" is revealed in the graphs as a dense semantic field of loss, betrayal, and intimate complaint; the "Environment" materializes in the lexicon of misery, the neighborhood, cigarette smoke, and marginal nocturnity; and the "Satirical" axis groups the vocabulary of moral judgment, the direct "I" versus "you", cynicism in the search for the "mango", the caricature of the upstart, and the lunfardesque edge. Crossing from one *cancionista* to another can feel, in some cases especially, like a change of lexical galaxy.

The comparative analysis immediately exposes a first unbridgeable fracture between two diametrically opposed ways of inhabiting the tango genre: the block of hegemonic modernist

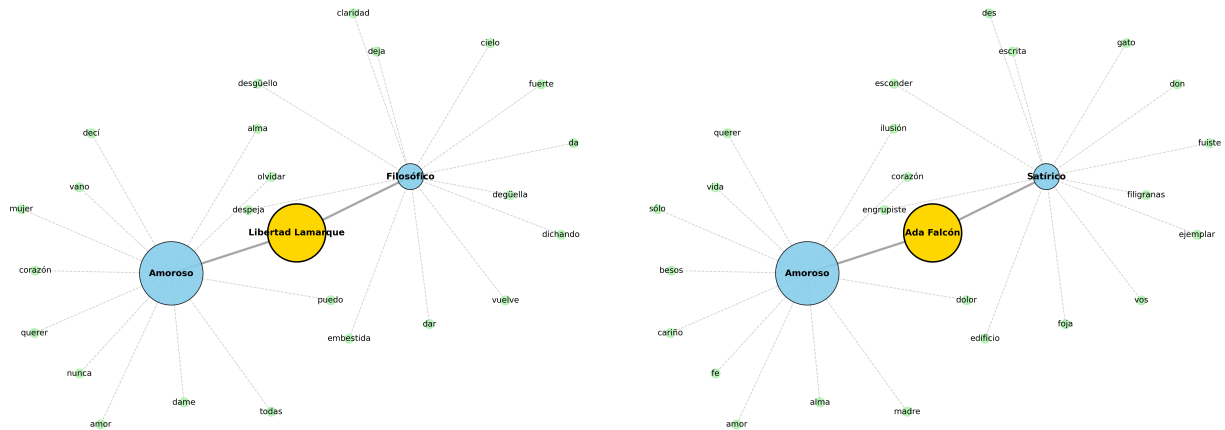


Figure 4. Archetypes of Sentimentalism: Libertad Lamarque and Ada Falcón. Force-directed graphs (TF-IDF). Both profiles show a massive dependence on the "Amorous" (Amoroso) node. The low number of lexical nodes (green dots) denotes a standardized vocabulary, centered on the romantic formula and purified of marks of orality or lunfardo. *Note: Key terms mapping for the original graph: Amoroso = Amorous.*

sentimentalism versus the subversion of popular speech that synthesizes the "lunfardesque" and the "orillero".

On the one hand, Libertad Lamarque, Mercedes Simone, and Ada Falcón make up the pole of amorous hegemony.

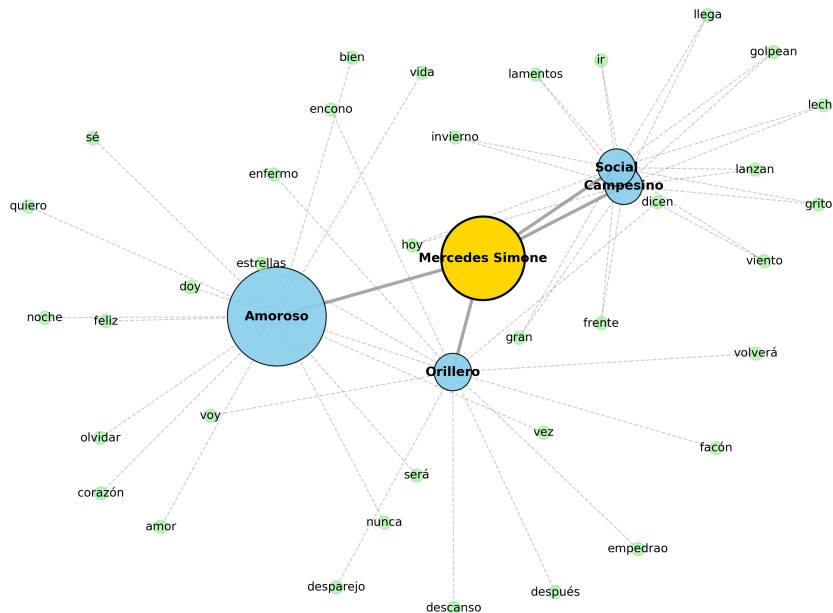


Figure 5. Mercedes Simone: The Stylization of Vocal Transvestism. The graph evidences the hegemony of the "Amorous" (Amoroso) axis anchored to the left (male/neutral voice zone). The minor satellites (Social, Orillero) suggest a slightly greater thematic openness than in Lamarque's model, although contained within cultured lyrics. *Note: Key terms mapping for the original graph: Amoroso = Amorous.*

escorted at the top by a significant Environment node. Merello does not sing of passive pain; her lexicon triggers irony, the *cachada*, and the raw chronicle of nocturnal life. Her Philosophical node, small and thrown to the far periphery of the map, suggests that her territory was not metaphysical abstraction, but the mud of experience.

In an aesthetically related, but much more expansive orbit, is located Rosita Quiroga. Although she shares with Merello the absolute dominance of the Satirical and the Environment (her largest nodes), Quiroga deploys a fragmented constellation that encompasses all the themes of the spectrum (Peasant, Metatango, *Orillero*, etc.). Her profile is deeply *arrabalero*, but of a total thematic voracity, touching every possible register from the matrix of the neighborhood.

Finally, the map reveals a third group characterized by intellectual complexity and existential tension: Tania, Sofía Bozán, and Azucena Maizani. Tania's constellation is an eloquent finding: her two largest nodes, Amorous and Philosophical, appear almost glued together in the left quadrant. This physical proximity in the graph recalls how, under the undeniable influence of Discépolo's repertoire, Tania fused affective complaint with ontological skepticism and reflection on time. Sofía Bozán, who in her theatrical performances would load her repertoire toward the "comic" side, who

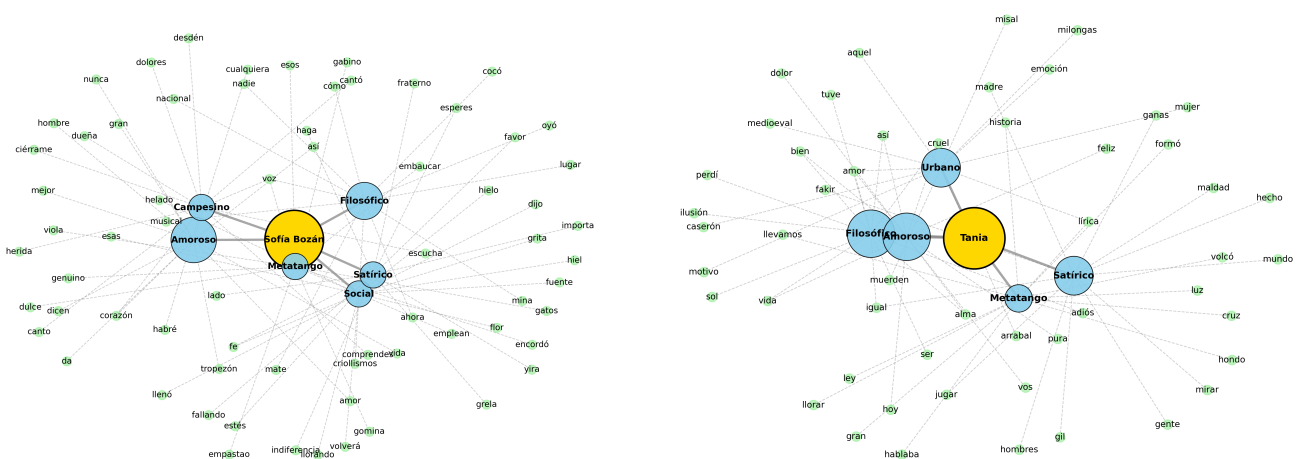


Figure 8. Ambiguity and Philosophy: Sofía Bozán and Tania. Profiles marked by intellectual tension. In Tania, the proximity of the "Amorous" (Amoroso) and "Philosophical" (Filosófico) nodes reflects the Discepolean influence. In Bozán, the dispersion among "Philosophical" (*Filosófico*), "Social", and "Environment" (*Ambiente*) reveals a versatility in record recording that transcends her unilateral stage image as a light vedette. *Note: Key terms mapping for the original graph: Amoroso = Amorous; Filosófico = Philosophical; Ambiente = Environment; Social = Social.*

went so far as to say "After I created the comic tango, I specialized in it and with it I went to Europe" (Stilman, 59), presents in her initial recordings a certain balance of forces (Amorous on the left, Philosophical on the right as great dynamos), but adding a counterweight, precisely, of the Satirical, which appears in this brief sample alongside exceptions — perhaps isolated — toward the Social and even the Peasant. Bozán is located in a place of central popularity that allowed her to premiere and explore the entire palette, even though her image has remained clearly fixed on the popular, comic, and defiant side.

Maizani, Sofía Bozán, or Rosita Quiroga saturate the visual space with an exuberance of lexical nodes. This proliferation responds empirically to their thematic dispersion: by simultaneously transiting through philosophy, the cabaret, satire, and lunfardo, these interpreters injected into the corpus a heterogeneous lexical richness, which the text mining algorithm rewards and deploys as a symptom of their discursive versatility.

In synthesis, text mining confirms that the label of "*cancionista*" should not operate historically as a reductionist umbrella. Far from uniformity, these women designed radically different repertoire engineerings: some established themselves as high priestesses of intimate pain, others as biting chroniclers of the urban margin, and a third group as interpreters of the philosophical anguish of River Plate modernity.

4.3. Topology of Enunciation: Orbits of Mimesis, Ambiguity, and Affirmation

The spatial projection of the corpus through a tri-polar gravity model then allows mapping how the interpreters and the thematic axes self-organize based on their statistical affinity toward the three lyrical voices (Neutral in the upper vertex, Male in the lower left, and Female in the lower right). This topological map suggests that the incursion of women in early tango did not form a uniform aesthetic block, but an ecosystem fractured by different strategies of generic negotiation.

In the exact center of the system emerges the figure of Ada Falcón, positioned equidistantly to the three poles and magnetically anchored by Theme 4 (Amorous), which occupies the heart of the graph. This location suggests that Falcón would represent the "balance point" or the central archetype of the genre at this stage, moving fluidly between the adoption of masks, the elision of the subject, and female enunciation, always pivoting on the axis of sentimentalism.

From this center of gravity, space bifurcates. Toward the right vector (Female Voice), the positioning of Libertad Lamarque stands out, who moves decidedly toward the affirmation of the woman's "I", but choosing almost always a more romantic language, more "adequate" to the expectations of female voice accepted by the tradition of modernist poetry. In the lower periphery of this same quadrant is located Tita Merello, whose orbit (strongly associated with the Environment and Satirical themes) suggests the construction of a subject that is also female, but very different: no longer the sentimental heroine, but the critical, ironic observer linked to the geography of the cabaret. Merello is the most radical expression of an uninhibited female presence, who appropriates an "in your face" popular language, and in that sense could be seen as the locomotive of the female appropriation of a space resistant to the classist establishment of the Buenos Aires of the twenties— which would manifest itself with the censorship of the 30s, precisely.

On the contrary, the left flank of the system evidences the historical weight of "vocal transvestism". A dense cluster formed by Tania, Azucena Maizani, and Rosita Quiroga groups in the center-left sector, while Sofía Bozán rises toward the Neutral vertex, although retained by the attraction of the Male orbit. Bozán, "queen of the cabaret" and a highly popular figure at that level of live tango, reveals nevertheless the circumscription of her voice to a male expectation that she had no problem

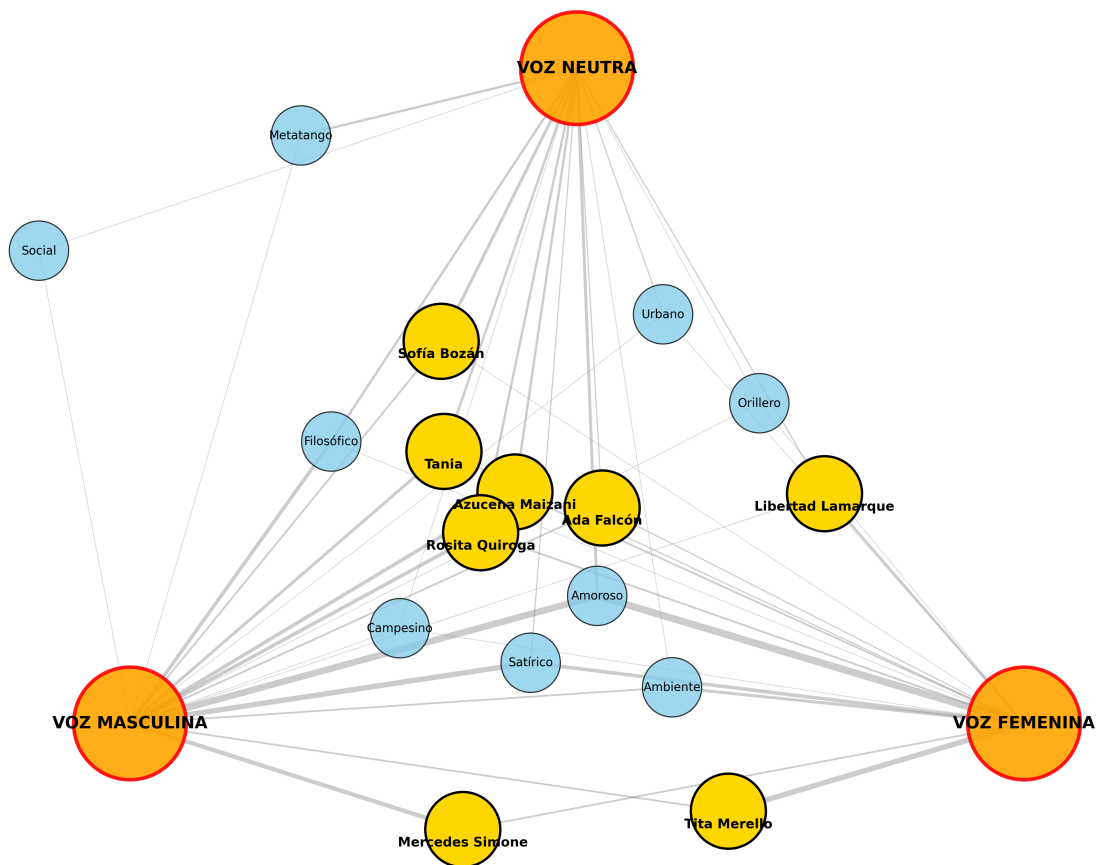


Figure 10. Tri-Polar Gravity Topology. Directed force map that locates the interpreters according to their statistical attraction toward the three poles of enunciation (Neutral, Male, Female). The graph exposes the fracture of the ecosystem: a cluster of male mimesis on the left, a vector of affirmation of female enunciation on the right (Lamarque, Merello), and positions of central equilibrium (Falcón). *Note: Key terms mapping for the original graph: Neutra = Neutral; Masculina = Male; Femenina = Female.*

incarnating. This constellation indicates that some of the pioneers cemented their legitimacy by relying on strategic ambiguity (Neutral Voice) or on the direct assumption of the voice of the *compadrito*.

Finally, the peripheral position of Mercedes Simone, strongly leaning toward the lower left extreme (Male Voice), turns out to be revealing when crossed with the thematic axes of that zone. The fact that themes of greater conceptual density such as the Philosophical, the Social, and the Metatango orbit exclusively in the left quadrant of the map suggests a strong bias of the era: existential reflection, collective denunciation, and tango metaliterature seemed to be discursive territories ontologically forbidden to the woman's voice. To sing about these topics, the *cancionista* saw herself obliged, almost inescapably, to invest the authority of the male voice.

The latter is nuanced when we observe that successful lyrics written by women like María Luisa Carnelli ("Se va la vida"), or even by the *cancionistas* themselves as is the case of Maizani ("Pero yo sé") or Simone ("Cantando"), adopt a radically female point of view, but repeat existing generic forms (the existential advice to the strayed man, or the romantic lament for the absent love). This

also shows why the "environment", "orillero", or "urban" categories gravitate toward the neutral voice, but with a certain female component, given by this type of interpretations.

4.4. Flows of Enunciation: The Ecosystem of Profiles through the Sankey Diagram

To summarize in a single image the internal dynamics of the corpus and the individual aesthetic decisions, a Sankey Diagram was finally implemented. This topological model allows visualizing the complete "ecosystem", tracing the flow of the repertoire from each *cancionista* toward her enunciation strategies (lyrical voice and lunfardo density) and, finally, toward her outlet in Vidart's thematic macro-axes. The reading of these flows reveals a marked specialization and, in parallel, strategies of deliberate versatility.

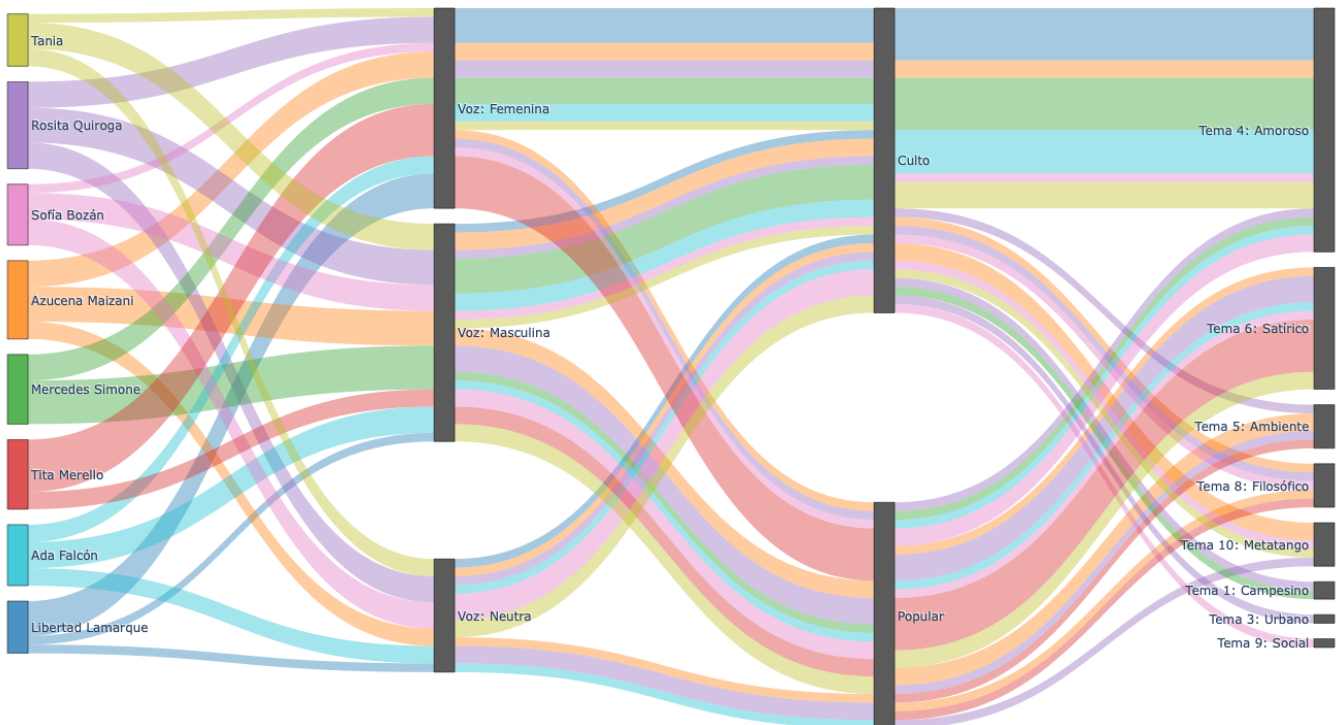


Figure 11. Flow Diagram (Sankey) of the *Cancionistas* Ecosystem 1922-1935. Modeling of the complete repertoire strategies. The flow is traced from the *cancionista* (left), passing through her Voice (Voz) and Language (Lenguaje) decisions, to the chosen Themes (Temas) (right). The thickness of the lines allows visualizing the specialization (e.g., Lamarque's univocal flow) versus the versatility (e.g., Quiroga's diversified flow). *Note: Key terms mapping for the original graph: Voz = Voice; Lenguaje = Language; Temas = Themes.*

At the extremes of the specialization spectrum, the graph exposes antagonistic positionings. On the one hand, Libertad Lamarque's flow is defined in an extreme and almost univocal way toward the Amorous axis (Theme 4), executed from a strict Female Voice. Lamarque thus consolidates the archetype of the sentimental heroine. In contrast, and sharing the same preference for the Amorous axis, Mercedes Simone's flow runs mostly toward the assumption of the Male Voice, sustaining her approach to sentimentalism from the practice of "vocal transvestism".

A third model of specialization is incarnated by Tita Merello. Her representative "river" falls definitely toward an affirmation of the female voice, but distanced from romantic passivity: her flow is strongly interwoven with the lunfardesque register and the axes of the *cachada*, configuring a defiant and plebeian female subject.

Faced with these bounded profiles, the diagram exposes figures of notable breadth. Rosita Quiroga emerges as the most comprehensive interpreter, diversifying her repertoire to cover practically all the thematic and enunciative spectrum available in the ecosystem. A comparable versatility is sustained by Azucena Maizani and Sofía Bozán; the latter transits strategically between the neutral and male voice, also broadly covering the *orillero* register.

Finally, the diagram sheds light on two philological phenomena of relevance. First, the sustained use of the Neutral Voice by Ada Falcón and Tania (in the case of the latter, predictably pulled by the interpretation of Enrique Santos Discépolo's repertoire, characterized by the elision of the subject and existential ambiguity). Second, the hierarchization of the densest topics of the corpus: the model reveals that it is Azucena Maizani and, to a lesser extent, Tania, who establish themselves as the undisputed referents of the most complex thematic areas, monopolizing the flows toward Theme 5 (Environment), Theme 8 (Philosophical), and Theme 10 (Metatango).

4.6 Quantitative synthesis of profiles

This heatmap summarizes and presents the absolute values of the corpus in terms of lyrical voice and type of language, per *cancionista*.

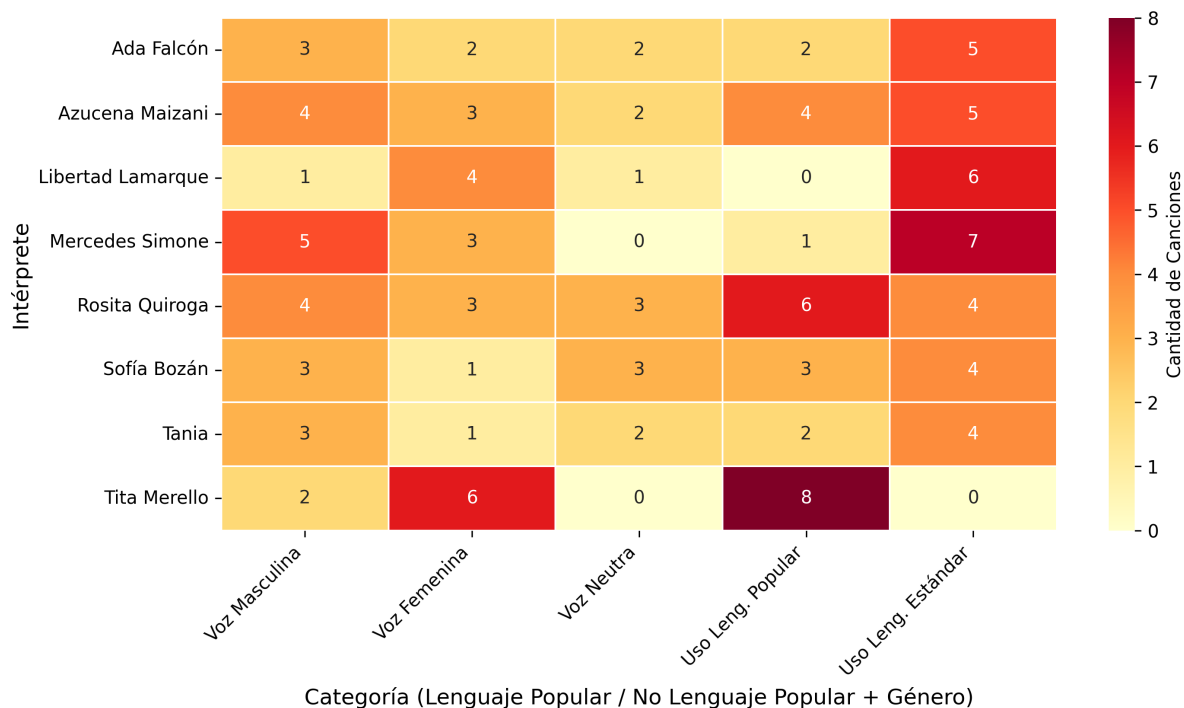


Figure 12. Heatmap: Matrix of Profiles and Intensity. Quantitative representation of the crossing of variables. The intensity of the color indicates the absolute frequency of songs per category, allowing a quick comparison of the volume of "Male Voice" (Voz Masculina) or use of "Popular Language" (Lenguaje Popular) among the different interpreters. *Note:* Key terms mapping for the original graph: Voz Masculina = Male Voice; Lenguaje Popular = Popular Language. 23

5. Discussion: Archetypes of the female voice in the foundational tango-canción?

By way of summary, we could already sketch some primary profiles of the considered *cancionistas*. Libertad Lamarque functions at this corpus level as an archetype of the "decent woman". Her profile moves away from the *arrabal*. Standard language and the female voice predominate. This reflects her melodramatic style, focused on the purist, romantic tango-canción stripped of street marginality—despite exceptions, such as her interpretation of "La Cumparsita".

At the other pole, Tita Merello uses the *orillero* and lunfardesque systematically in all her interpretations, and there is in her—even more significant due to the combination—a marked dominance of the female voice in the vast majority of her representative songs. Tita Merello, both in her image and in her language, recounts her own reality without "cultured" filters, representing a sort of denunciation of the hypocrisy and manners imposed in society.

Mercedes Simone, often considered the technically most gifted singer of the group, has a peculiar profile: she almost does not use lunfardo (except in one song, "Estampa rea") and manages to distribute her voice between male "transvestism" and female emission, in lyrics composed by men (Manuel Ferradás Campos and Luis Rubistein), or by herself. From a formal image, notably married to a musician who sometimes accompanied her on stage, she was able to adopt the male "I" or the female "I", in lyrics written by men or by herself, from an elegant and salon aesthetic, far from the extremes of the "género chico" and the revue.

Azucena Maizani and Rosita Quiroga represent in this group better than the rest a sort of versatile and pioneer vein. They present the most balanced and varied profiles. They navigate between lunfardo and standard Spanish, and assume female, male, and neutral roles almost equally. Tania shows something that could be described as a certain strategy of neutrality. Instead of defining herself like Merello or Lamarque, she opted for a generic and thematic ambiguity that largely comes from her artistic association with Enrique Santos Discépolo. She thus provided, perhaps, the female public with a pathway into Discépolo's disenchanted philosophy. Using neutral or "genderless" lyrics could operate in her as a way of appropriating tangos written for men without having to force the voice toward a specific identity.

Sofía Bozán, on her part, the most popular "vedette" on stage, alternates non-*orillero* and *orillero* language. A figure of the strongest theatrical company of the moment, the Muiño-Alippi, she was often called upon for the premiere role of pieces that would become classics, such as "Yira Yira" or "Un Tropezón". She herself developed very early the uninhibitedness of popular lexicon and humor on stage—in that, she is a precursor to Tita Merello, who would use it well later—. But Bozán's successes in cinema or on record do not always reflect that vein that characterized her on stage. It is she, for example and in direct distancing from her public archetype, who opens the film *Tango!* (1933) representing a country girl with "Mi provincianita". Later, precisely that film will represent her passage from the mud hut to the theatrical stage, in a simple synthesis of the creators' self-awareness regarding the sociology of the tango phenomenon.

6. Conclusions

The computational and philological analysis of the foundational corpus of the *cancionistas* considered in the period (1922-1935) allows disarticulating the notion of a homogeneous or passive female block within the history of tango. Far from being limited to replicating without friction a repertoire forged by and for men, the eight analyzed interpreters deployed markedly divergent enunciation, lexical selection, and thematic appropriation strategies.

In the first place, the statistical crossing of the lyrical voice demonstrates that the installation of the woman in tango was a space of active and complex negotiation. Although "vocal transvestism"—the mimetic assumption of the male perspective and pronoun—represented the relative majority strategy (41%), the consolidation of a female lyrical "I"—with the relativizations regarding its representativeness in relation to any notion of a "genuine" female voice—reached a convincing 37.7%. This narrow margin, added to the use of ambiguity (Neutral Voice, 21.3%) by figures like Tania or Ada Falcón, evidences that male enunciation was not an unbreakable mandate, but an aesthetic code that the interpreters compromised, subverted, or eluded according to their own identity constructions.

Secondly, the text mining applied to Vidart's taxonomy empirically exposes the falsity of any stereotype that could historically confine the *cancionista* to sentimental lament. Although the Amorous Theme dominates the ecosystem at a macro level (46%), the extraction of lexical constellations reveals that the tango "popular language" was inhabited by female voices, from singular archetypes. Interpreters like Libertad Lamarque and Mercedes Simone incarnated the hegemony of intimate pain through cultured language, although often female—and even, written by herself in the case of Simone—; at the antipodes, Tita Merello and Rosita Quiroga dynamited said canon appropriating satire (23% of the global corpus), lunfardesque irony, and the raw landscape of the *arrabal*. Even if it is accepted that "*the meaning of tango discourse is macho*" (Dos Santos, 2021, p. 11), and that the representations of women in that lyrical universe would be reduced to the trio "*milonguera [evil], girlfriend [good], and mother [ideal or summum of good]*" (Varela, 2005, pp. 83-84), the corpus and its analysis suggest, nevertheless, that the *cancionistas* did not limit themselves to repeating that repertoire even when they sang it (they also modified it and added texts of their own authorship), but that the relevance of their performances, their internal differentiation that generates distinct individualized profiles, and their character as popular stars and talented artists generated a pole of power and female presence not controllable that intervenes and initiates the transformation of that hegemonic male universe. Simultaneously, repertoire architectures of the spectacle in vogue, like those of Azucena Maizani and Sofía Bozán, demonstrated that the female voice was fully capable of sustaining and massifying topics of high conceptual density, monopolizing the flows toward philosophical tango, metadiscursive reflection, and the environment painting of the cabaret.

Ultimately, the distant reading of this ecosystem proves that the label of "*cancionista*" should not function historiographically as a reductionist umbrella or confirmatory of overly general theses. Through the sharp contrast in grammatical manipulation, the use of popular language, and thematic selection, these pioneers operated as vehicles of a hegemonic discourse, and at the same time as true

creators of their image and their discourse. With their decisions and versatilities, they not only guaranteed the commercial viability of women in an industry dominated by men, but rewrote from within the semantic boundaries of the tango-canción, leaving a multifaceted legacy that is barely beginning to be mapped with precision.

Notes

¹ Declaration on the Use of Artificial Intelligence (AI Acknowledgement): The authors declare having used the generative model Gemini 3.1 Pro during the period of February and March 2026. The use of this tool was limited to the following instrumental purposes: (a) assistance in the writing and debugging of the source code in the Python language implemented for text processing (TF-IDF) and network analysis; (b) generation of the graphic visualization scripts (Sankey, Sunburst, heat maps, and Force-Directed Graphs) from the qualitative and quantitative data extracted and curated exclusively by the authors; and (c) assistance in style correction, translation, and formatting of bibliographic references. Artificial intelligence was not used at any time to generate or alter the corpus data, nor to formulate the interpretive conclusions of the study. The authors assume full responsibility for the functionality, accuracy, and replicability of the codes and visualizations presented herein.

² It is significant that "El porteño" (with two versions that are precisely the first and the last recording by Vivianne), consisted of lyrics rewritten by Ángel Villoldo in the female gender: "*Soy hija de Buenos Aires / por apodo la criollita / la moza más compadrita / que en este suelo pisó / cuando un tango en la vigüela / rasguea algún compañero / no hay nadie en el barrio entero / que baile mejor que yo*" [I am a daughter of Buenos Aires / by nickname the *criollita* / the most *compadrita* lass / who stepped on this ground / when a tango on the *vigüela* / some companion strums / there is no one in the whole neighborhood / who dances better than me]. This adaptation of lyrics to female voices is not the only one. Among the first tango lyrics, even within the period prior to the one that concerns us, "La Morocha" stands out, in 1905 ("*Yo soy la morocha / la más agraciada / la más renombrada / de esta población...*" [I am the *morocha* / the most graceful / the most renowned / of this town...]). Both follow the same strategy: the woman defines herself in the text, thus creating a female space of enunciation already at the dawn of the genre.

³ We have included a brief biographical sketch of each of them in Appendix 2.

⁴ Marcela Gené (2012) has argued, thematizing the figure of the "Sisebuta," or the dominant and physically violent wife very popular at the time, that the inversion of gender roles and female authority could be deployed freely within the frameworks of humor, which would explain the space that the satirical and costumbrist occupy as one of the dominant poles of the female repertoire. Quantitatively, and independently, our research tends to confirm that observation. This does not detract from the fact that the presence of women like Merello or Quiroga in that register also represented a challenge to the morality and class expectations that were established regarding women.

⁵ Perhaps the case of Anita Palmero, or that of Herminia Velich, could be considered as a ninth member, but their work, permanence, and transcendence do not justify their inclusion, beyond their historical value. Estela Dos Santos (2021), who has written one of the few studies dedicated to the *cancionistas*, defines an initial "pantheon" that includes Lamarque, Tania, Falcón, Simone, Quiroga, Maizani, Merello. Gobello (2002) in his complete collection of biographical sketches includes our eight, and they are the only ones who enter for the considered period. There are specialized biographies dedicated to Merello (Delgado, 2006); Lamarque, Simone, Quiroga, among others, have received diverse biographical treatment and very numerous reviews and biographical sketches that it would be impossible to mention. Ada Falcón was the subject of a documentary (Wolff, 2003), Tita Merello of another, Lamarque wrote a successful autobiography, etc. In general, there is a solid consensus regarding the impact and visibility of these eight female names within the considered period.

⁶ The other analysis we know of with a quantitative approach is that of Darío Cantón, which, independently of its virtues, only takes into account 5 songs with a female lyrical voice, and bases its entire analysis of that variant on them. See Cantón (1972), pp. 22 - 52. In general, the compilations of tango lyrics and poets, which abound, do not take into account the gender factor ("tangos sung by women") to populate their anthologies.

⁷ In terms of massification through the theater, Mazziotti (1985, pp. 75-76) reports the opening of 60 theaters in downtown Buenos Aires alone between 1880 and 1930. If in 1911 there were 21 theaters, by 1928 they already totaled 43. Regarding the number of spectators, by 1925 a high average of 2.6 attendances per year per inhabitant was recorded, in a city of almost 7 million inhabitants.

⁸ *"In 1932, only two films were shot [still silent films]. The following year, there were six. In 1935, the number grew to thirteen, reaching twenty-eight in 1937 and fifty in 1939, closing the decade with nine studios and about thirty companies that employed almost four thousand people and two thousand five hundred theaters"* (Gil Mariño, 2023, p. 67).

⁹ See the detailed classification in Appendix 1.

¹⁰ María Luisa Carnelli is an important author. Dorner Linne (2009, pp. 143-144) registers and names 44 songs only between 1927 and 1929, of which only two ("El malevo" and "Dos lunares (Nadie sabe)") were signed as "Mario Castro." Among those are tangos of great repercussion such as "Cuando llora la milonga" (with Juan de Dios Filiberto), "Pa'l cambalache" (with Rafael Rossi), "Se va la vida" (with Edgardo Donato), and "El malevo" (with Julio de Caro), among others. Carnelli recalls her experience composing "Se va la vida," and a brief example of the rejection of a typical male exponent of the "official" tango imaginary toward her lyrics: *"Around 1929 with Edgardo Donato I composed 'Se va la vida'. As at that time almost all tangos had lunfardo lyrics and I already knew that language through my brothers and the journalistic environment I frequented, I had to use it. On the occasion of a tribute to Filiberto, my colleague Last Reason [Reference to Máximo Sáenz (1886-1960), a Uruguayan turf and costumbrist chronicler of impact in the River Plate press during the considered years] sat next to me. At that meeting, the guest orchestra played 'Se va la vida'. Then Reason commented that it was a really beautiful tango, except for the lyrics. He freely criticized it and finally I told him it was mine and you should have seen the awkward moment I put him through. That tango became popular in Spain brought by Azucena Maizani. It was very well liked there, to such an extent that Manuel Pizarro commented to me that the most requested tangos were 'La cumparsita', 'Adiós muchachos' and 'Se va la vida'."* Pínsón, Néstor, "María Luisa Carnelli". In Pínsón (n.d.-b).

References

Assunção, F. (1984). *El tango y sus circunstancias*. Editorial El Ateneo.

Barsky, J. y O. (2012). *La Buenos Aires de Gardel*. Sudamericana.

Barsky, J. (2016). *El tango y las instituciones: de olvidos, censuras y reivindicaciones*. Teseo.

Cantón, D. (1972). *Gardel, ¿a quién le cantás?*. De la Flor.

Cortázar, Augusto Raúl (1976). *Ciencia folklórica aplicada*. Buenos Aires, Fondo Nacional de las Artes.

Delgado, J. (2006). *Tita Merello, la morocha argentina*. Aguilar.

Dorner Linne, A. (2009) *Poesía y tango*. Ediciones El Tranvía.

Dos Santos, E. (1998). *Las mujeres del tango*. Corregidor.

García Jiménez, F. (1964) *El tango, Historia de medio siglo (1880-1930)*. Eudeba.

Gené, M. (2012). 'Sisebutas' en Buenos Aires. Family strips de los años veinte. *Antíteses*, 5(9), 33–58. <https://doi.org/10.5433/1984-3356.2012v5n9p33>

Gil Lozano, F. (2008). Género y representaciones femeninas en el cine sonoro argentino. En María Celia Bravo (Ed), *Historia de luchas, resistencias y representaciones. Mujeres en la Argentina, siglo XIX y XX* (pp. 391-406). Edun.

Gil Mariño, C. (2023). De percantas, otarios y guapos a la broadcasting. Ficciones de género y de la nación en las primeras décadas del cine argentino. En *Tango que fuiste y serás* (pp. 64-73). Biblioteca Nacional Mariano Moreno.

Gobello, J. (2002). *Mujeres y hombres que hicieron el tango*. Centro Editor de Cultura Argentina S.R.L.

Gobello, J. (1976). Orígenes de la letra de tango. En *La historia del tango. Los orígenes* (pp. 99-144). Corregidor.

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

- Mazziotti, N. (1985). El auge de las revistas teatrales argentinas. En *Cuadernos Hispanoamericanos*, Num. 425 (pp. 73-75).
- Mazzucchelli, A. (2006). El modernismo en el tango. *Revista de Crítica Literaria Latinoamericana*, 32(63/64), 11–28. <https://doi.org/10.2307/25070322>
- Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.
- Pellettieri, O. (2011). El sainete criollo: teatro de autor y de actor. En Mirza (Ed.), *Teatro y representación. Perspectivas contemporáneas sobre teoría, historia y crítica del teatro latinoamericano y europeo* (pp. 37-42). CSIC-UdelaR.
- Pinsón, N. (s.f.-a). *¿Quién fue la primera mujer que grabó tangos?* TodoTango. Recuperado el 1 de marzo de 2026 de <https://www.todotango.com/historias/cronica/373/Quien-fue-la-primera-mujer-que-grabo-tangos/>
- Pinsón, N. (s.f.-b). *María Luisa Carnelli*. TodoTango. Recuperado el 1 de marzo de 2026, de <https://www.todotango.com/creadores/biografia/855/Maria-Luisa-Carnelli/>
- Rossi, Vicente (1958). *Cosas de negros*. 2a ed. Librería Hachette.
- Saikin, M. (2004) *Tango y género. Identidades y roles sexuales en el tango argentino*. Abrazos.
- Stilman, E. (1965). *Historia del tango*. Editorial Brújula.
- Tallón, Juan Sebastián (1964). *El tango en sus etapas de música prohibida*, 2a ed. Instituto Amigos del Libro Argentino.
- UC Santa Barbara Library. (s.f.). *Discography of American Historical Recordings (DAHR)*. Recuperado entre el 4 de febrero - 5 de marzo de 2026, de <https://adp.library.ucsb.edu/>
- Underwood, T. (2018). The Transformation of Gender in English-Language Fiction. *Journal of Cultural Analytics*, 3(2), 1-25. <https://doi.org/10.22148/16.019>
- Varela, G. (2005). *Mal de tango: historia y genealogía moral de la música ciudadana*. Paidós.
- Vidart, Daniel (1967). *El tango y su mundo*. Ediciones Tauro.
- Vilariño, Idea (1965). *Las letras de tango*. Ed. Schapire S.R.L..

Wolf, S., & Muñoz, L. (Directores). (2003). *Yo no sé qué me han hecho tus ojos* [Película]. Habitación 1520 Producciones; Instituto Nacional de Cine y Artes Audiovisuales (INCAA)

Appendices

Appendix 1: Daniel Vidart's Taxonomy (Vidart, 1965, pp. 96-98)

1. The *peasant* theme a) Landscapes and characters of rural life; b) the erotic world; c) the *criollo* duel; d) the philosophy and knowledge of folk.
2. The *orillero* theme a) The suburban presence: the outskirts, the neighborhood, the *arrabal*; b) the evocation of the neighborhood; picturesque characters (the Italian in tango and in sociodemographic reality, etc.); d) the *bailongo*; e) *orillera* and *arrabalera* typology: *tuitas*, *paicas*, *lunfas*; f) the *orillera* heraldry: the *funyi*, the dagger, the lions, the little lantern, the patio of the *conventillo*, the street corners, the *boliche*, the *bulín*, the *catrera*, etc.
3. The *urban* theme a) Streets of the city; b) events of the city; c) human archetypes of the city; d) the port and its world (the docks, the canteens, the characters); e) the melancholic evocation of the city.
4. The *amorous* theme (the sentimental tango) a) Faithful love: the mother, the girlfriend, the "sweet home", the children. b) The sorrow of love. c) The betrayal of love (attitudes towards the *espiente*): a) betrayal and revenge; b) betrayal and forgiveness; c) betrayal and nostalgia; d) betrayal and rejoicing; e) betrayal and alcoholism; d) Seduction; e) Abandonment; f) Love and friendship.
5. The *environment* a) Female typology: a) the *milonguero* boom and the brevity of glory; b) the decline of the gallant woman and its aftermath. b) Male typology: the *compadrito* dancer, the *muchacho rana*, the womanizer, the *canfli*, the *bacán*, the *otario*, the *miché*, the *caralisa*, the fallen *guapo*, the old age of the "*calavera*". c) Description of the milonga and its false tinsel. e) Vices: alcohol, narcotics, etc.
6. The *satirical* theme (the *cachada* tango) a) Critique of customs; b) critique of people and character caricatures; c) critique of vital situations.
7. The *playful* theme a) The prototypes: the *escolaseador*, the *burrero*, the *quinielero*, etc. b) the *timba*: cards, races, lotteries, roulette, sports: football, boxing.
8. The *philosophical* theme (the *Weltanschauung* tango) a) General philosophy of life and its psychological concomitants: a) bitterness, sadness, pain; b) cynicism; c) skill and *cancha*; d) aggressiveness; e) *guaranguería*; f) spite; g) disillusionment; h) indifference; i) oblivion; j) loneliness. b) Attitudes towards death. c) Valorization of the present: *carpe diem* and everyday axiology. d) Evocation of the golden yesterday: nostalgia, recreation, return. e) Presence and absence of God.
9. The *social* theme a) Exploited and exploiters; b) misery and hunger; c) the strike; d) the *gayola*.

10. *Tango* as a theme

Appendix 2

Brief biographical sketch of the selected *cancionistas*

Rosita Quiroga started in the theater, where she remained for years, until her friendship with the wealthy Villar Sáenz Peña family opened the doors to the Victor label (Gobello, 2002, p. 266), where she debuted in 1922. Being the first of the *cancionistas* of this stage to reach the record, she opened, along with Maizani (Dos Santos, 2011, p. 41), the path for the rest, and became a central figure, although with her own style of interpretation, with a close, confidential voice, and a strong *arrabalero* accent. "*She was the counter-image of Azucena, of Ada, of Mercedes, of la Liber*" (Gobello, 2002, p. 266). She stopped recording in the 1930s, perhaps harmed by political changes. She sang on the radio from 1924 to 1930, and then her voice was also eclipsed for many years in that medium. (Dos Santos, 2011, 40).

Sofía Bozán began as a chorus girl in the Vittone-Pomar and Muiño-Alippi company, and premiered her first tango on stage ("Canillita") in 1925. She was above all a great theatrical star, with less transcendence in radio and cinema, and not too many recordings (Gobello, 2002, 96).

Azucena Maizani had been a radio voice, "*milonguera* in the Tabaris" (Gobello, 2002, 253) and chorus girl. She debuts as her own personality, with a strong, direct and intense voice, "torn, declamatory" (Dos Santos, 2011, p. 41), after being "discovered" as a singer by Enrique Delfino. Her first resounding success was the interpretation, within a *sainete*, of the tango "Padrenuestro", in July 1923. She immediately recorded her first tango with the Victor company. In the considered period she was the one with the most impact—and the most profitable—although she died in poverty. "*Azucena and Gardel created the way of singing tango*" (Gobello, 2002, p. 254). She was also the one who most visually blended in with the typical male figures (she often appears disguised as a "*malevo*" and a gaucho), "but not transvestite, but as a fully grown woman who had dressed as a man" (Gobello, 2002, p. 253).

Mercedes Simone did not start out in the theater, but singing live, accompanied by her husband, the guitarist Pablo Rodríguez. She had a very brief stint at the Maipo theater, but Simone "*only wanted to sing. She did not go on stage to gesticulate or shake her body*" (Dos Santos, 2021, p. 47). She reached the record thanks to Rosita Quiroga, who took her to Victor to record her first song in December 1927. Her great medium of diffusion and triumph was the radio. There is broad agreement in considering her the technically best of all the *cancionistas*, and probably the one who inspired a more distant and, at the same time, lasting respect.

Tita Merello was already part of the "Bataclán" theater (Dos Santos, 2021, p. 26) when she had not yet turned 16, and a *cancionista* on stage in the early 1920s, where she covered all genres: comedy, *sainete* and *género chico* (Dos Santos, 2021, p. 27). By 1924 she had reached the Maipo theater, where she premiered or sang numerous tangos. Of a long life, career and presence in all River Plate media, the cinema is, of all the media she employed, the one that ended up identifying her the most.

Ada Falcón, who had been a *cupletista* in her adolescence and early youth, was "discovered" by Enrique Delfino in 1929, who took her to the record, where she recorded just over 200 songs. She generally withdrew from direct contact with the public. She had an intense, long and complex romance with the director Francisco Canaro between 1928 and 1938 (Dos Santos, 2021, p. 35).

Following the emotional crisis linked to the breakup with Canaro, she decided on an early retirement in 1942, and lived from then until her death isolated and secluded in a town in the province of Córdoba. "It could be said that Ada Falcón was exclusively a radio and record star" (Gobello, 2002, p. 22).

Libertad Lamarque also began in the theater, with the Carcavallo company, where she acted and sang. In 1930 she debuted in silent cinema. "*Her career is a succession of triumphs in radio, on records, on stage, on sets*" (Gobello, 2002, p. 15), both in Argentina and abroad. Her career was very long and she continued acting and triumphing everywhere almost until her death, at 92 years old. She left the Buenos Aires scene for a long time after a personal clash with Eva Perón in 1945.

Tania, Spanish by birth, also first entered the theatrical scene in Morocco, along with her first husband Antonio Fernández and his group "Los Mexican". She arrived in Buenos Aires in 1924, and after a contact with guitarist Mario Pardo, she entered the cabaret "Chantecler" where she began singing tangos—including Discépolo's "Esta noche me emborracho". José Razzano heard her sing and connected her with Discépolo around 1927. Tania and the composer began—from the very day they met, according to Tania (Gobello, 2002, p. 100; Dos Santos, 2021, p. 49)—a long amorous relationship, and a career of collaboration, where Tania recorded many of Discépolo's tangos, and also premiered some of them.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.