



Análisis literario computacional de letras de tango: métrica, emociones y evolución temática

Computational Literary Analysis of Tango Lyrics: Meter, Emotions, and Thematic Evolution

Ricardo Martínez-Gamboa

Universidad Diego Portales

<https://orcid.org/0000-0002-5348-469X>

ricardo.martinez@udp.cl

Recibido: 21/10/2025

Aceptado: 12/11/2025

Resumen: Este estudio interdisciplinario se propone investigar 6.811 letras de tangos escritas en el rango temporal que va desde 1912 hasta 1985, organizadas por letristas, periodos y generaciones, mediante un análisis literario computacional, mapeando su evolución métrica, emocional y temática. Se utilizó SICAM para la métrica automatizada, NES para el análisis de sentimientos y LIWC y LDA TF-IDF para el modelado de tópicos. Los hallazgos métricos indican que en las primeras etapas del tango dominaron los versos de arte menor, especialmente el octosílabo, reflejando su carácter festivo. Posteriormente, aumentaron las estructuras complejas y el versolibrismo, debido a la influencia de la modernización urbana. Se confirmó una tendencia sostenida hacia una mayor negatividad emocional, lo que apoya la tesis de Borges de que el tango “va languideciendo y entristeciéndose”. Temáticamente, a lo largo de su historia, aumentaron los tópicos que denominamos *emociones*, *muerte* (asociada a la pérdida del “Buenos Aires querido”) y *trabajo*. La disminución en el tiempo del tópico *persona* subraya un aumento progresivo de la abstracción y un menor enfoque narrativo en las letras. Además, el uso de lunfardismos decae luego de 1930. Este estudio ofrece una visión cuantificada de cómo las transformaciones estéticas del tango contribuyen a construir las complejidades de la modernidad rioplatense.

Palabras clave: tango, análisis computacional, métrica poética, emociones, evolución temática.

Abstract: This interdisciplinary study aims to investigate 6,811 tango lyrics written between 1912 and 1985, organized by lyricist, period, and generation, using computational literary analysis to map their metrical, emotional, and thematic evolution. SICAM was used for automated metrics, NES for sentiment analysis, and LIWC and LDA TF-IDF for topic modeling. The metrical findings indicate that in the early stages of tango, short verses, especially the octosyllable, dominated, reflecting its festive character. Subsequently, complex structures and free verse increased due to the influence of urban modernization. A sustained trend toward greater emotional negativity was confirmed, supporting Borges's thesis that tango “is languishing and growing sadder.” Thematically, throughout its history, the topics we call emotions, death (associated with the loss of “beloved Buenos Aires”), and work increased. The decline over time in the personal theme underscores a progressive increase in abstraction and a less narrative focus in the lyrics. Furthermore, the use of lunfardo (Argentine slang) declined after 1930. This study offers a quantified view of how the aesthetic transformations of tango contribute to constructing the complexities of modernity in the Río de la Plata region.

Keywords: Tango, Computational Analysis, Poetic Metrics, Emotions, Thematic Evolution.



Los contenidos de este artículo están bajo la licencia de Creative Commons Reconocimiento - No Comercial - Compartir Igual 4.0 Internacional.

El tango, como género musical y expresión cultural, ha sido históricamente un vehículo privilegiado para la exploración de los afectos y las dinámicas de la modernidad urbana. Esto ha sido documentado profusamente, entre otros, por Vilariño (1981, 1995) quien, luego de su indagación cualitativa, sostiene que el tango puede tratar temas asociados al dolor, la nostalgia, la ternura, la resignación, la piedad, la amargura, la impotencia, la tristeza, así como múltiples formas del amor y del odio; o por Romano (2007), quien asegura que las letras del género reflejan en un alto porcentaje los deseos, sufrimientos, alegrías y frustraciones del porteño, lo que lo convierte en una escuela del sentimiento nacional¹.

Nacido en los márgenes de Buenos Aires y Montevideo a finales del siglo XIX, el tango emergió como una síntesis de las tensiones sociales, económicas y culturales que marcaron la transición de estas ciudades hacia la modernidad (Godio y Mancuso 2006). En sus letras, melodías y coreografías, el tango encapsula un universo de emociones que van desde la nostalgia y el desamor hasta la crítica social y la melancolía, lo que lo convierte en un espejo de las transformaciones urbanas y las pasiones humanas que las acompañaron, elaboradas mediante propiedades poéticas singulares (García 2008b; Gobello 1967; Martínez-Gamboa 2019; Mina 2006, Ulla, 1982).

La relación entre el tango y los afectos ha sido ampliamente documentada en la literatura académica. Investigadores como José Gobello (1967) y Ricardo García (2008b) han mostrado cómo las letras de los tangos de la Guardia Vieja y la Guardia Nueva reflejan las experiencias de los sectores populares en un contexto de migración, urbanización y cambio social. Desde los arrabales y conventillos hasta los cafés y cabarets, el tango se convirtió en una crónica poética de la vida urbana, dando voz a los sentimientos de pérdida, añoranza y resistencia que caracterizaron a las clases trabajadoras y a las y los inmigrantes que poblaron las ciudades del Río de la Plata. Tangos como “El Choclo” (1903) de Ángel Villoldo y “La cumparsita” (1916) (García 2008a) de Gerardo Matos Rodríguez ejemplifican esta conexión entre las letras –agregadas posteriormente–, la música y las transformaciones urbanas, aludiendo tanto a los espacios físicos como a las emociones asociadas a ellos. Sin embargo, el análisis sistemático de las letras del tango desde una perspectiva cuantitativa y computacional sigue siendo un terreno poco explorado.

En este contexto, el presente estudio propone analizar las letras de los tangos utilizando diversas técnicas, que incluyen: a) el procesamiento automatizado de su métrica, b) el análisis de sentimientos (*sentiment analysis*), a través del método de Procesamiento del Lenguaje Natural (NLP, sigla en inglés) de análisis de textos que permite identificar, extraer, cuantificar y estudiar sistemáticamente los estados afectivos y la información subjetiva que portan los enunciados lingüísticos, y c) el *topic modeling*, metodología de análisis de textos que permite identificar patrones temáticos latentes en grandes corpus de datos. Estos métodos se han utilizado con éxito en disciplinas como la literatura, la historia y la sociología para explorar las propiedades formales de los textos, así como las preocupaciones, valores y emociones de la época o comunidad que los escriben. Aplicados al tango, estos enfoques ofrecen la posibilidad de mapear las principales temáticas que atraviesan sus letras y de rastrear cómo evolucionaron a lo largo del tiempo, desde los primeros tangos de la Guardia Vieja hasta las innovaciones de Astor Piazzolla y el electrotango del siglo XXI (Liska 2008).

El objetivo de este trabajo es doble. En primer lugar, buscamos identificar los principales recursos métricos poéticos, los sentimientos y los tópicos temáticos presentes en las letras del tango y analizar cómo estos se relacionan con los afectos y las dinámicas de la

¹ Este estudio se ha realizado bajo el proyecto ANID Fondecyt Iniciación N°11230032 (2023-2025).

modernidad urbana. En segundo lugar, pretendemos explorar cómo las transformaciones históricas y culturales del tango se reflejan en la evolución de sus temáticas, considerando las distintas etapas del género: la Guardia Vieja, la Guardia Nueva, las grandes orquestas de los años cuarenta y cincuenta, y las innovaciones de Piazzolla y el electrotango. Este enfoque nos permitirá no solo comprender mejor el lugar del tango en la historia cultural del Río de la Plata, sino también contribuir al estudio más amplio de la canción popular como expresión de las pasiones humanas en contextos urbanos determinados.

La importancia de este estudio radica en su capacidad de ofrecer una perspectiva novedosa sobre el tango a través de la combinación de herramientas computacionales con un enfoque cultural y emocional. Mientras que los estudios tradicionales sobre el tango han tendido a centrarse en análisis cualitativos de letras individuales o en la historia de sus principales exponentes, el uso de métodos como el análisis métrico automatizado, el análisis de sentimientos y el *topic modeling* nos permiten abordar el género desde una perspectiva más amplia y sistemática. Al analizar un corpus representativo de letras de tango, podemos identificar patrones estilísticos y temáticos que trascienden las obras individuales y revelan las preocupaciones y emociones colectivas de las comunidades que dieron vida a este género.

El tango, como expresión de la modernidad urbana, está profundamente arraigado en las tensiones y contradicciones de la vida en la ciudad. En las letras de los tangos de la Guardia Vieja, por ejemplo, encontramos una mezcla de nostalgia por la vida rural perdida y fascinación por las oportunidades y peligros de la ciudad (García 2008b; Gobello 1967). Tangos como “El entrerriano” (1897) de Rosendo Mendizábal y “La morocha” (1905) de Enrique Saborido y Ángel Villoldo capturan esta dualidad, combinando imágenes de los arrabales y los conventillos con referencias a la vida nocturna y el ascenso social. Letras de tangos como “Mi noche triste” (1917) de Pascual Contursi y Samuel Castriota y “Mano a Mano” (1923) de Celedonio Flores (Monjeau 1998) reflejan un cambio hacia una mayor introspección emocional, explorando temas como el desamor, la traición y la melancolía. En las décadas siguientes, el tango evolucionó hacia formas más complejas y sofisticadas, con las grandes orquestas de los años cuarenta y las innovaciones de Piazzolla en los años sesenta y setenta, lo que llevó al género a nuevos niveles de experimentación musical y poética (Liska 2008; Mina 2006; Navarro 2008).

El análisis de las letras de tango a través de los tres métodos computacionales señalados nos permitirá rastrear cómo estas temáticas evolucionaron a lo largo del tiempo y cómo se relacionan con los cambios sociales, económicos y culturales de los habitantes del Río de la Plata. Por ejemplo, podemos explorar el modo en que la nostalgia por el pasado rural y la crítica a la modernidad urbana se transforman en las letras de las grandes orquestas, como sucede en “Tinta roja” (1941) de Cátulo Castillo y Sebastián Piana, o la manera en que las innovaciones de Piazzolla reflejan un alejamiento de los tópicos tradicionales del tango para hacer una exploración más abstracta y universal de los afectos, como se ve en “Balada para un loco” (1969) o “Libertango” (1974), aunque en este caso ello se produce por la composición musical, dado que se trata de una pieza instrumental, a pesar de que posteriormente algunas personas le agregaron letra. Asimismo, el *topic modeling* nos permitirá identificar patrones temáticos menos evidentes, como la representación de la masculinidad, la relación entre el tango y la política, o la influencia de la inmigración en las letras del género (Espíndola 2002; Sarlo 1995).

En términos metodológicos, este estudio se levanta desde un corpus representativo de

6.811 letras de tango, o géneros afines, extraídas de la base de datos de Rosati², pobladas a su vez desde los repositorios Todo Tango³ y El Recodo Tango⁴ que abarcan las principales etapas del género. Utilizamos técnicas de procesamiento de lenguaje natural para limpiar y preparar los textos, y aplicamos algoritmos de *topic modeling* como Latent Dirichlet Allocation (LDA) para identificar los tópicos temáticos latentes en el corpus, así como el análisis de sentimientos, *Bag-of-Words*, clasificación métrica de estrofas/stanzas y detección de lunfardismos. Los resultados fueron analizados e interpretados en relación con el contexto histórico y cultural del tango, a través de un enfoque interdisciplinario que combina herramientas computacionales con perspectivas de la historia cultural, la sociología y los estudios literarios. Debe considerarse, con todo, que la recabación de data de letras de tangos desde bases de datos como estas adolecen de dos problemas. De una parte no son transparentes los criterios ecdóticos –de la edición de textos– que se han seguido en el poblamiento de la base, y de otro, no hay detalle en ellas de la versión del tango que se considera. Aun así, para los objetivos del presente estudio, se puede continuar con la data de la que se dispone, y se puede proponer que en estudios sucesivos es posible afinar estos aspectos, toda vez que los objetivos de esta investigación son posibles de ser desarrollados incluso con estas consideraciones.

En conclusión, buscamos contribuir al estudio de la lírica tanguera a través de una perspectiva innovadora sobre la relación entre afectos, modernidad urbana y canción popular en el tango. Al analizar las letras del tango utilizando estas técnicas, esperamos no solo arrojar nueva luz sobre las dinámicas emocionales y culturales del género, sino también abrir nuevas vías para el estudio de la canción popular como expresión de las pasiones humanas en contextos urbanos. En un momento en que el tango sigue siendo una fuente de inspiración y reflexión para artistas y académicos (González y Rolle 2004), este enfoque nos permitirá redescubrir su riqueza temática y su relevancia para comprender las complejidades de la vida moderna.

Metodología y procesos realizados

Este estudio se inspira en el trabajo de Germán Rosati (2022) titulado “Procesamiento de Lenguaje Natural aplicado a las ciencias sociales. Detección de tópicos en letras de tango”, que constituye un modelo metodológico innovador para el análisis computacional de textos en humanidades. Dicho artículo aplica técnicas de Procesamiento de Lenguaje Natural (NLP) y modelado de tópicos, específicamente Latent Dirichlet Allocation (LDA), para identificar temas recurrentes en un corpus de 5.617 letras de tango y, a partir de tal enfoque, logra una sistematización y escalabilidad en el análisis de datos textuales que supera las limitaciones de los métodos cualitativos tradicionales.

Con esta investigación perseguimos avanzar respecto al modelo propuesto por Rosati mediante la ampliación del corpus de letras de tango, incorporando nuevas fuentes para enriquecer la base de datos y mitigar posibles sesgos derivados de la selección inicial, así como la adición de dos técnicas más de análisis, a saber, el análisis de sentimientos y el análisis métrico automatizado de las letras. Además, se explora de manera específica el uso del lunfardo, analizando su frecuencia, contexto y relación con los tópicos detectados. Este

² https://github.com/gefero/topic_modeling_tango/tree/master

³ <https://www.todotango.com/>

⁴ <https://www.el-recodo.com>

enfoque permite profundizar en las particularidades lingüísticas del tango.

Otro aspecto innovador del presente estudio es la cronologización de la evolución de los tópicos, las emociones y las estructuras métricas, así como sus combinaciones, tomando como referencia tanto las fechas de publicación de cada uno de los 6.811 tangos como las fechas de nacimiento de los letristas para establecer generaciones de autores. Este análisis permite mostrar vínculos entre las temáticas predominantes y los contextos históricos y generacionales de los compositores, ofreciendo una perspectiva más detallada sobre los cambios en los temas, emociones y estructuras métricas abordados y desplegados en el tango a lo largo del tiempo.

Se ha utilizado el modelo de *bolsas de palabras* (*Bag-of-Words*, BoW) en dos aplicaciones diferentes, la primera ocupando las listas de palabras LIWC, y la segunda a través de un levantamiento manual realizado por un par de codificadoras expertas, de una lista clasificada de palabras clave, con la finalidad de investigar temáticas específicas dentro del corpus, como las emociones, las referencias al arrabal y las imágenes urbanas. Este enfoque ha permitido identificar patrones y matices en el uso del lenguaje, alimentados desde el algoritmo de Latent Dirichlet Allocation (LDA).

En conjunto, buscamos no solo replicar el modelo metodológico de Rosati, sino también expandirlo y adaptarlo para abordar preguntas más específicas sobre el tango como fenómeno cultural y lingüístico. Este enfoque interdisciplinario espera contribuir al entendimiento de las “pasiones cantadas” en el tango, destacando su riqueza temática y su evolución histórica.

En las siguientes secciones metodológicas se explica el levantamiento del corpus, así como cada uno de los tres procedimientos que se utilizaron para estudiar los tangos. A saber, el análisis de sentimientos, el analizador métrico (SICAM) y el *topic modeling*, que usará en primer lugar la bolsa de palabras LIWC, y en segundo lugar una lista de palabras significativas en los tangos determinada por LDA y luego etiquetada manualmente.

El corpus de tangos se ha extraído de la base de datos de German Rosati, actualizada respecto de la última versión que este investigador argentino utilizó para su publicación de 2022. La base actualmente contiene 6.811 obras musicales que se distribuyen de la siguiente manera:

Tabla 1
Distribución de registros por género musical

Género	Frecuencia	Porcentaje
Tango	5741	84.3
Vals	637	9.4
Milonga	333	4.9
Otros	100	1.5
Total	6811	100.0

Para el presente estudio se ha considerado la colección completa, entendiendo que todas estas obras corresponden al estilo tanguístico y sus géneros afines, más allá de sus diferencias formales. Debe señalarse en este aspecto, para profundizarlo, que en ocasiones es posible que la data entre en conflicto como base de datos con los corpus canónicos de ciertos letristas. Es este último, por ejemplo, el caso de las letras de Enrique Santos Discépolo consignadas por Rosati, toda vez que canónicamente el letrista realizó los textos de solo

veintisiete tangos, mientras que en la base hay noventa y cinco letras. Estas diferencias para el presente estudio no han sido delimitadas, por lo que debe trabajarse posteriormente en este asunto.

La base de datos de Rosati contiene la siguiente información que ha sido extraída para la base de este estudio:

Tabla 2

Campo	Descripción
LETRA	Letra (lyrics) del tango
FECHA	Fecha de publicación del tango (aaaa-mm-dd)
AUTOR	Letrista del tango
AUTOR2	Segundo letrista del tango (de haberlo)
COMPOSITOR	Autor de la música del tango
DIRECTOR	Director de la orquesta del tango (de estar registrado)
ORQUESTA	Nombre de la orquesta del tango (de estar registrada)
CANTANTE	Nombre del cantante del tango (de estar registrado)
SOLISTA	Nombre del solista del tango (de estar registrado)
TÍTULO	Nombre del tango
TOTAL	10 variables

A esta data inicial se han agregado, luego de llevar a cabo los procesos de análisis métrico, análisis de sentimientos y *topic modeling*, otros elementos que se señalarán luego de detallar cómo se han llevado a cabo estos procesos.

Para realizar los procedimientos sobre el corpus se ha usado SICAM⁵ (Silabizador del Castellano Moderno; Martínez-Gamboa *et al.* 2025), un sistema computacional diseñado para el análisis métrico automatizado de poesía en español, desarrollado principalmente en Perl, con módulos complementarios en Python. Su objetivo es realizar un análisis detallado y preciso de sílabas, versos, estrofas y poemas completos, integrando reglas fonotácticas del español y técnicas de *machine learning*. Este sistema es especialmente útil para abordar las complejidades métricas de la poesía hispana, incluyendo licencias poéticas y estructuras flexibles.

SICAM opera mediante ocho módulos principales que trabajan en serie. El proceso comienza con la silabificación, que divide las palabras en sílabas considerando reglas fonotácticas y licencias poéticas como la sinalefa, la diéresis y la sinéresis. Luego, el módulo de acentuación identifica las sílabas tónicas y átonas, y ajusta el análisis según las modificaciones métricas. El módulo de rima analiza la última palabra de cada verso para determinar el tipo de rima (consonante, asonante o libre) y su disposición en la estrofa. Posteriormente, el módulo de clasificación de estrofas combina los resultados anteriores para identificar el tipo de estructura estrófica, utilizando una base de datos de métricas conocidas y un modelo de *machine learning* basado en el algoritmo Random Forest, entrenado con 3.503 estrofas clasificadas manualmente.

El sistema permite realizar tareas como la separación en sílabas de palabras individuales, el conteo de sílabas gramaticales y métricas en versos, la identificación del axis rítmico (ritmos trocaicos y yámbicos), la clasificación de versos según la tradición poética

⁵ <https://www.tecling.com/cgi-bin/sicam/index.pl>

(por ejemplo, endecasílabo heroico), y la determinación del tipo de rima y estrofa. SICAM también ofrece un análisis detallado del poema como unidad completa, integrando los resultados de cada módulo. El *output* del programa incluye tablas con información detallada sobre la silabización, acentuación, tipo de verso, rima y métrica, junto con la clasificación estrófica basada en reglas y *machine learning*. Además, presenta una tabla de rimas que muestra las relaciones entre los versos de la estrofa.

SICAM puede ejecutarse como programa independiente mediante intérpretes de Python y Perl o en un servidor web usando la librería CGI de Perl. Su diseño modular y enfoque híbrido lo convierten en una herramienta robusta y flexible para el análisis literario computacional, ideal para estudiar la métrica de tangos y otras formas poéticas en español (ver Figura 1).

El cálculo de la polaridad emocional se realizó utilizando el modelo NES, basado en el sistema desarrollado por Lage (2014) para el análisis de polaridades en Twitter. Este modelo clasifica las emociones de los textos en tres categorías: positivas (+1), negativas (-1) y neutras (0). La base del análisis es un lexicón etiquetado manualmente que asigna valores emocionales a palabras clave. Este lexicón puede ser ampliado mediante procedimientos automáticos y asegurar una cobertura más amplia del vocabulario.

El proceso incluye tres pasos principales: 1) identificación de palabras clave en el texto mediante el lexicón etiquetado, 2) análisis de estructuras gramaticales que puedan alterar la polaridad, y 3) clasificación final del texto según su carga emocional. Este enfoque permite calcular la polaridad de unidades textuales más extensas, como versos, manteniendo un alto nivel de precisión.

La validación del modelo se ha realizado mediante el coeficiente Kappa, comparando los resultados automáticos con un etiquetado manual realizado por evaluadoras expertas. Este procedimiento asegura la fiabilidad del análisis, especialmente en textos líricos, donde la interpretación emocional puede ser más subjetiva. La metodología es adaptable a otros géneros literarios, como en este caso los tangos, considerando las particularidades lingüísticas y emocionales de cada corpus. El análisis de polaridad se llevó a cabo sobre los tangos individuales, que alcanzan en el corpus la cifra de 6.811 casos (ver Figura 2).

Para el proceso de *topic modeling*, se emplearon dos enfoques metodológicos principales: LIWC, y LDA y TF-IDF combinados, cada uno con características específicas que permiten extraer y analizar información lingüística y temática de los textos. LIWC⁶ es una herramienta de análisis de texto que utiliza un diccionario de palabras y raíces organizadas en categorías lingüísticas, emocionales y cognitivas. Este método analiza los textos palabra por palabra, y asigna cada término a una o más categorías predefinidas. El procesamiento incluye el cálculo del porcentaje de palabras que coinciden con las categorías del diccionario, lo que genera resultados que pueden ser exportados a programas estadísticos como SPSS o Excel. LIWC permite identificar patrones emocionales, cognitivos y temáticos en los textos, lo que proporciona una visión estructurada de los componentes lingüísticos. En este estudio, se utilizó para analizar las emociones y procesos cognitivos presentes en los tangos, así como para identificar temas como la amistad, la religión o la sexualidad.

⁶ <https://www.liwc.app/>

Figura 1. Ejemplo de visualización de la salida de la aplicación de SICAM para la milonga “Jacinto Chiclana” (1965), letra de Jorge Luis Borges y música de Astor Piazzolla.

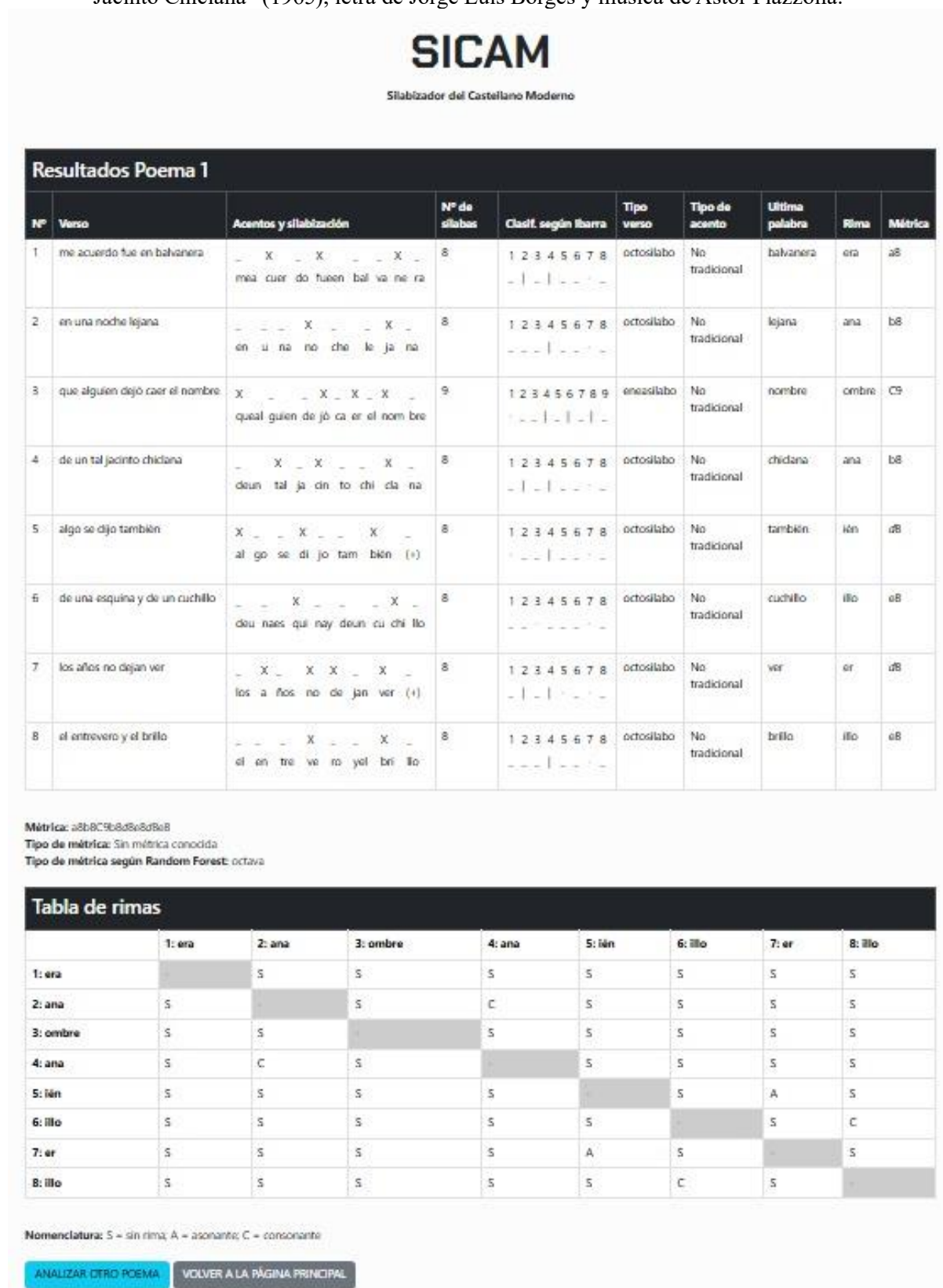
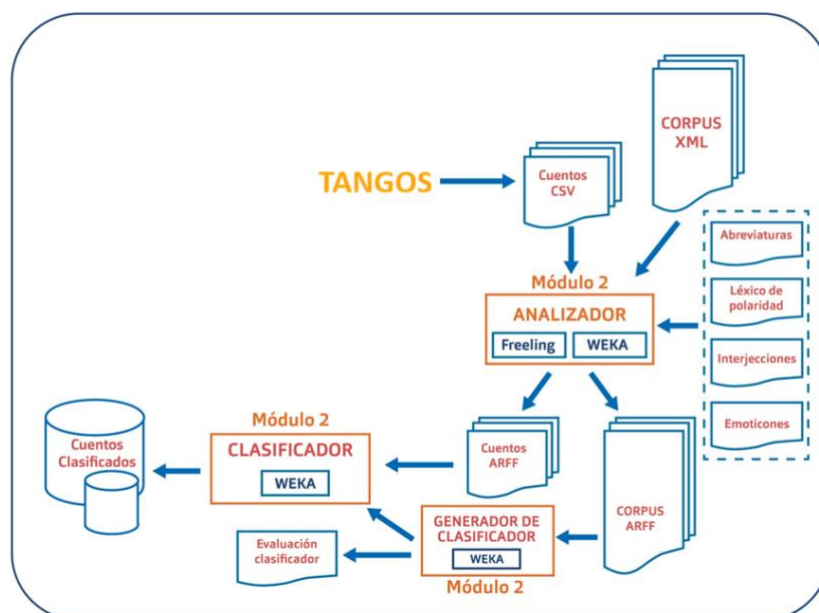


Figura 2. Algoritmo de Lage (2014).



El análisis LIWC ocupó el diccionario Spanish 2007 (Ramírez-Esparza *et al.* 2007) que consta de 12.551 palabras (lemas) categorizadas manualmente en 74 categorías que incluyen, por ejemplo, los siguientes rótulos en inglés o mediante siglas: *social, familia, amigos, humanos, afect, EmoPos, EmoNeg, ansiedad, enfado, triste, MecCog, insight, causa*, etcétera. Debe señalarse que los nombres de estos rótulos en alguna medida están inspirados en nombres análogos de rótulos para el inglés, por lo que en ocasiones se mantienen en la versión para el castellano de esas denominaciones. El análisis de LIWC se llevó a cabo sobre cada uno de los 6.811 tangos como unidades independientes.

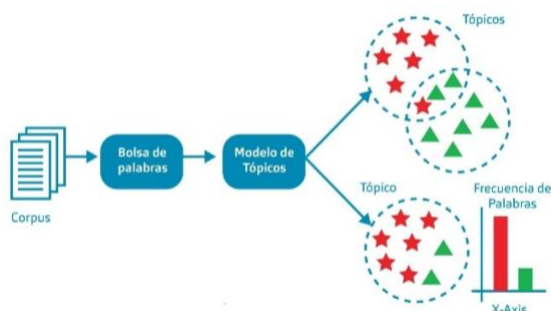
LDA (Jelodar *et al.* 2019) es un modelo generativo que asume que cada texto es una mezcla de varios tópicos y que cada palabra pertenece a uno de estos. Dicho enfoque utiliza el método de bolsa de palabras (BoW), donde el orden de las palabras no afecta el análisis, lo que permite identificar patrones temáticos recurrentes en un corpus. LDA es especialmente útil para descubrir tópicos latentes en grandes conjuntos de textos, como los tangos, y para generar modelos temáticos que reflejen las relaciones entre palabras y categorías. En este estudio, LDA se aplicó para identificar los temas predominantes en los tangos, como el amor, la nostalgia o la vida urbana, y para analizar cómo estos tópicos se distribuyen en el corpus.

Se generaron por medio de LDA 20 listados o bolsas con 50 palabras cada uno para el corpus, considerando todos los 6.811 tangos de la base de datos, los que posteriormente fueron etiquetados temáticamente por codificadoras expertas siguiendo un riguroso protocolo de clasificación, y obteniendo como resultado las siguientes categorías: *acción, lunfardo, emoción, tiempo, música, naturaleza, objeto, religión, persona, flora, fauna, lugar, s/c* (sin clasificación).

TF-IDF (Nikolenko *et al.* 2017) es una medida numérica que evalúa la relevancia de una palabra en un documento dentro de un corpus. Calcula la frecuencia de término (TF) y la frecuencia inversa de documento (IDF), y les asigna mayor peso a las palabras que son significativas en un texto, pero menos comunes en el conjunto de documentos. Este método es útil para identificar palabras clave y filtrar términos triviales como artículos o

preposiciones. En este estudio, TF-IDF se utilizó para extraer los términos más representativos de cada tango, como apoyo al LDA, y así facilitar el análisis de las palabras que caracterizan los textos y su relación con los tópicos identificados. Estas metodologías se complementaron para ofrecer un análisis integral de los tangos desde perspectivas lingüísticas y temáticas (ver figura 3).

Figura 3. Ilustración (Atkinson-Abutridy 2022) de los resultados del *topic modeling*.



Para determinar la presencia de lunfardismos en las letras de tango del corpus, se implementó un procedimiento sistemático que consistió en los siguientes pasos: a) se extrajeron todos los lemas (formas básicas de las palabras, como “amar” para “amo” o “amábamos”) del diccionario de lunfardo de Todo Tango⁷, que contiene alrededor de 15.000 entradas. Se exploró, asimismo, en los textos de Conde (2011), que indica que el lunfardo corresponde a un repertorio léxico compuesto por voces y expresiones de diversos orígenes, utilizadas en alternancia con el español estándar, así como de Gobello (1991) y Espíndola (2002). Posteriormente, se alineó y contrastó esta lista con el leuario de Freeling, que constituye el repositorio más extenso del español con alrededor de 670.000 entradas. Este contraste permitió identificar términos exclusivos del lunfardo que no tienen homónimos ni acepciones distintas en el español general, denominados “lunfardismos puros”.

Este enfoque metodológico, común en la elaboración de diccionarios referenciales (Morales Pettorino 1984), asegura una mayor precisión en la identificación de vocablos específicos del lunfardo. Por ejemplo, el lema “ligar” en el diccionario de lunfardo se define a través de múltiples acepciones relacionadas con la suerte, el castigo físico y el juego, mientras que en el *Diccionario de la Real Academia Española* (DLE) se limita a la acción de “atar” y, en México, es una forma de referirse a la seducción, la coquetería, el cortejo erótico: “ligarse a alguien” es seducirlx / “estar ligando” es coquetería mutua. Este análisis computacional se centró exclusivamente en los lunfardismos puros, evitando ambigüedades derivadas de términos con significados compartidos entre el lunfardo y el español general.

La base de datos generada finalmente para el corpus se organiza de la considerando las fechas de nacimiento de los letristas y una división de los periodos de las obras en estadios de cinco años (1921-1925, 1926-1930, por ejemplo), y las generaciones a las que pertenecen los letristas (1890-1904, 1905-1919, etc., siguiendo el modelo de Coupland 2001):

⁷ <https://www.todotango.com/comunidad/lunfardo>

Tabla 3

Descripción de los campos de la base de datos (verso y metadatos)

Campo	Descripción
ID	Número identificador del tango
Letra	Letra (lyrics) del tango
Fecha	Fecha de publicación del tango (aaaa-mm-dd)
Año	Año de publicación del tango
Periodo	Periodo de publicación del tango (1921–1925, 1926–1930...)
Autor	Letrista del tango
Año nacimiento	Año de nacimiento del letrista del tango
Generación	Generación de nacimiento del letrista del tango (1890–1904, 1905–1919...)
Autor 2	Segundo letrista del tango (de haberlo)
Compositor	Autor de la música del tango
Director	Director de la orquesta del tango (de estar registrado)
Orquesta	Nombre de la orquesta del tango (de estar registrado)
Cantante	Nombre del cantante del tango (de estar registrado)
Solista	Nombre del solista del tango (de estar registrado)
Título	Nombre del tango
LIWC	Categorías según BoW-LIWC (Amigos, Ansiedad, Certeza...)
Polaridad	Polaridad (Análisis de Sentimientos) del tango (+1, -1, 0)
Acentos	Ritmo acentual del verso (--+---+-)
Tipo de verso	Métrica del verso (octosílabo, endecasílabo...)
Estrofa por verso	Tipo de estrofa en que se incluye el verso (cuarteto, décima...)
Lunfardismos	Número de palabras del lunfardo que ocurren en cada tango
TOTAL	21 variables

Resultados y discusión

Luego del levantamiento del corpus y de los análisis de la data se obtuvieron los resultados que se presentan a continuación. Los autores principales de las letras de los tangos y géneros afines son los siguientes:

Tabla 4

Campo	n
Letristas (únicos)	877
Tangos (total)	6799

Letrista	n
Enrique Cadícamo	289
Homero Manzi	201
Carlos Bahr	167
Francisco García Jiménez	154
Celedonio Flores	153
José María Contursi	146
Juan Andrés Caruso	120
Manuel Romero	117
Homero Expósito	105
Cátulo Castillo	101
Luis Rubinstein	98
Enrique Santos Discepolo	94
Héctor Marcó	85
Pascual Contursi	83
Horacio Sanguinetti	82
Ivo Pelay	78
Reinaldo Yiso	77
Emilio Fresedo	76
José González Castillo	72
Alfredo Le Pera	70

El número de tangos letrados por generación de nacimiento de los autores – considerando que solo se ha encontrado la fecha de nacimiento de 579 letristas– según el modelo de Coupland (1991) es el siguiente:

Tabla 5
Distribución de tangos por generación del letrista (frecuencia y %)

Generación	n (tangos)	Porcentaje
1860	6	0,6 %
1875	84	8,7 %
1890	305	31,7 %
1905	336	34,9 %
1920	204	21,2 %
1935	28	2,9 %
Total	963	100,0 %

Nota. Porcentajes calculados sobre el total; la fila Total corresponde a la suma.

Es notorio que las tres generaciones con más letristas identificables por su fecha de nacimiento son las de 1890, 1905 y 1920, lo que corresponde epocalmente a la Tardo Guardia Vieja, la Guardia Nueva y la era de las Grandes Orquestas. Para solidificar el listado de 6.811 tangos por fecha se periodizaron sus años de publicación en estadios de cinco años (1921-1925 o 1936-1940, por ejemplo) y se llegó al siguiente resultado:

Tabla 6

Distribución de tangos por período (Inicio–Fin): frecuencia y porcentaje

Inicio	Fin	n (tangos)	Porcentaje
1911	1915	1	0,0 %
1916	1920	14	0,2 %
1921	1925	87	1,3 %
1926	1930	1.047	15,4 %
1931	1935	525	7,7 %
1936	1940	545	8,0 %
1941	1945	1.313	19,3 %
1946	1950	766	11,2 %
1951	1955	878	12,9 %
1956	1960	699	10,3 %
1961	1965	389	5,7 %
1966	1970	299	4,4 %
1971	1975	103	1,5 %
1976	1980	92	1,4 %
1981	1985	53	0,8 %
Total		6.811	100,0 %

Nota. Porcentajes calculados sobre el total de 6.811 tangos; la fila Total corresponde a la suma.

Se puede observar en la tabla que los dos periodos más prolíficos son los de la Media Guardia Nueva (1926-1930), en especial el momento más fructífero de los discos de Gardel y la era de las Grandes Orquestas (1941-1945), esto es, el periodo posgardeliano en que emergen las orquestas que siguen el modelo imperante en Norteamérica y el Caribe de una orquesta multinstrumental más un *crooner*. El estudio de la evolución de la polaridad (valores promediados que pueden ser de +1, -1 o 0 para cada tango) arroja por periodo los siguientes resultados:

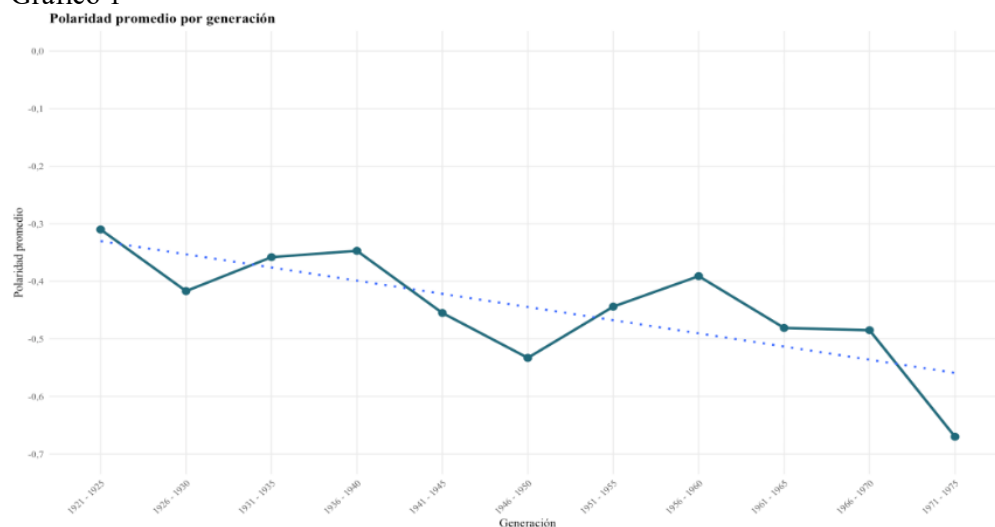
Tabla 7

Polaridad promedio por generación (análisis de sentimientos)

Generación	Polaridad promedio
1921 - 1925	-0,310
1926 - 1930	-0,417
1931 - 1935	-0,358
1936 - 1940	-0,347
1941 - 1945	-0,455
1946 - 1950	-0,533
1951 - 1955	-0,444
1956 - 1960	-0,391
1961 - 1965	-0,481
1966 - 1970	-0,485
1971 - 1975	-0,670

Nota. Valores de polaridad promediados por período; formato con coma decimal.

Gráfico 1



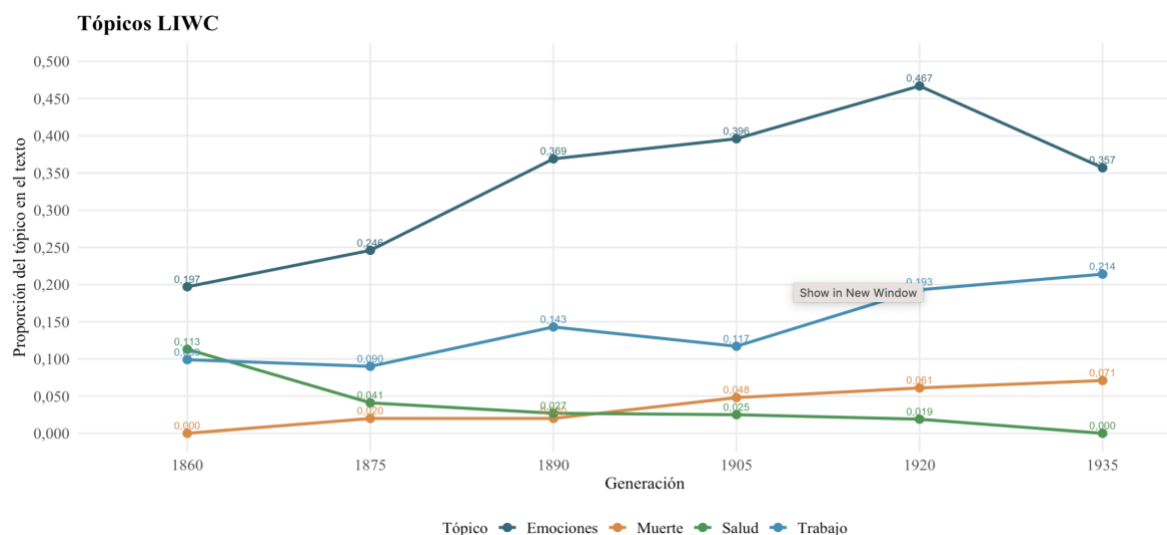
Se puede observar en el Gráfico 1 una clara tendencia a aumentar la negatividad de las letras de los tangos en la medida en que transcurre el tiempo. Esto refrenda el hallazgo de Rosati (2022), quien ya daba cuenta de ello en su estudio. Rosati se apoya en la tesis de Jorge Luis Borges, quien hacia la década del 1960 planteaba que “el tango, como hemos visto, [cuando] empezó, surge de la milonga, y es al principio un baile valeroso y feliz. Y luego, el tango va languideciendo y entristeciéndose” (2016: 80-81). Los tópicos LIWC detectados se presentan en Tabla 8 a continuación:

Top LIWC por generación de nacimiento del autor/a

Top LIWC	1860	1875	1890	1905	1920	1935
Amigos	0,000	0,041	0,051	0,022	0,024	0,000
Ansiedad	0,000	0,027	0,029	0,054	0,052	0,000
Certeza	0,056	0,018	0,041	0,054	0,066	0,000
Conflicto	0,000	0,028	0,037	0,033	0,057	0,000
Cuerpo	0,211	0,081	0,140	0,140	0,222	0,143
Dinero	0,028	0,052	0,067	0,053	0,052	0,000
Emociones	0,197	0,246	0,369	0,396	0,467	0,357
Enojo	0,000	0,033	0,074	0,088	0,113	0,000
Entendimiento	0,099	0,208	0,175	0,192	0,156	0,214
Esparcimiento	0,282	0,118	0,158	0,231	0,189	0,071
Familia	0,042	0,032	0,038	0,029	0,057	0,000
Hogar	0,028	0,058	0,082	0,074	0,047	0,000
Incertidumbre	0,070	0,073	0,073	0,086	0,080	0,000
Inhibición	0,042	0,028	0,026	0,029	0,000	0,000
Logro	0,042	0,066	0,131	0,125	0,184	0,000
Muerte	0,000	0,020	0,020	0,048	0,061	0,071
Personas	0,197	0,066	0,060	0,048	0,066	0,071
Reflexión	0,268	0,373	0,392	0,409	0,519	0,286
Relaciones sociales	0,282	0,242	0,277	0,247	0,358	0,286
Religión	0,042	0,087	0,082	0,109	0,123	0,071
Salud	0,113	0,041	0,027	0,025	0,019	0,000
Sexualidad	0,070	0,036	0,044	0,060	0,061	0,071
Trabajo	0,099	0,090	0,143	0,117	0,193	0,214
Tristeza	0,014	0,026	0,068	0,104	0,146	0,000

Nota. Valores LIWC promediados por generación; separador decimal con coma.

Gráfico 2



Aquí podemos observar que el único tópico que disminuye es el de *salud*, mientras que los otros tres aumentan. Esta disminución se puede relacionar con el desarrollo de la medicina en el Buenos Aires y el Montevideo que ingresan al siglo XX. En ese contexto, los servicios hospitalarios y de medicamentos se hicieron más eficientes y científicos, lo que generó una disminución de males como las enfermedades, que pasan a ser menos relevantes en la vida diaria (Puppo, Nóbile y Sader 2022).

Contrariamente, el tópico *muerte* va ligeramente aumentando, probablemente en un sentido metafórico, asociado a la pérdida del “Buenos Aires querido” (De Majo 1998) –lo que se refuerza con el aumento de la mención de las *emociones*–. Lo mismo sucede en el caso de *trabajo*, en la medida en que el progreso de las sociedades rioplatenses disminuye los oficios esporádicos, y con *cesantía*, la cual se asocia a la emergencia del aparato y los funcionarios públicos. Además de lo anterior, y ahora periodizados cada cinco años, se pueden observar otras evoluciones de tópicos LIWC en los siguientes tres gráficos:

Gráfico 3

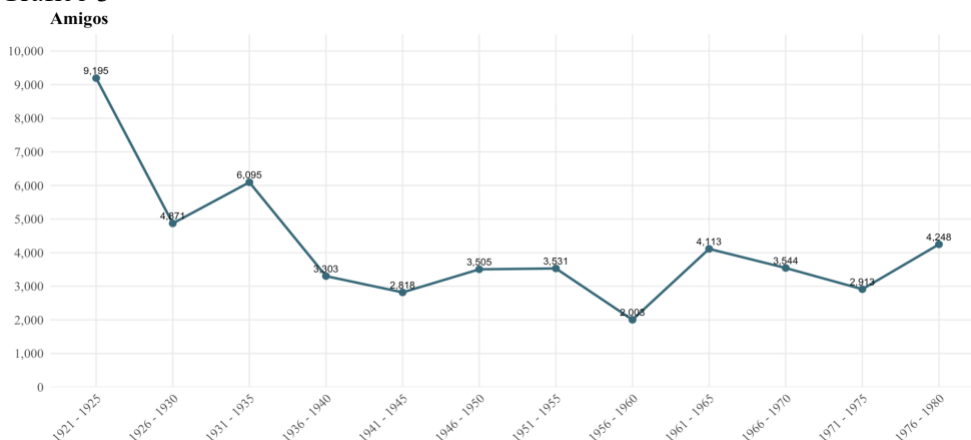


Gráfico 4

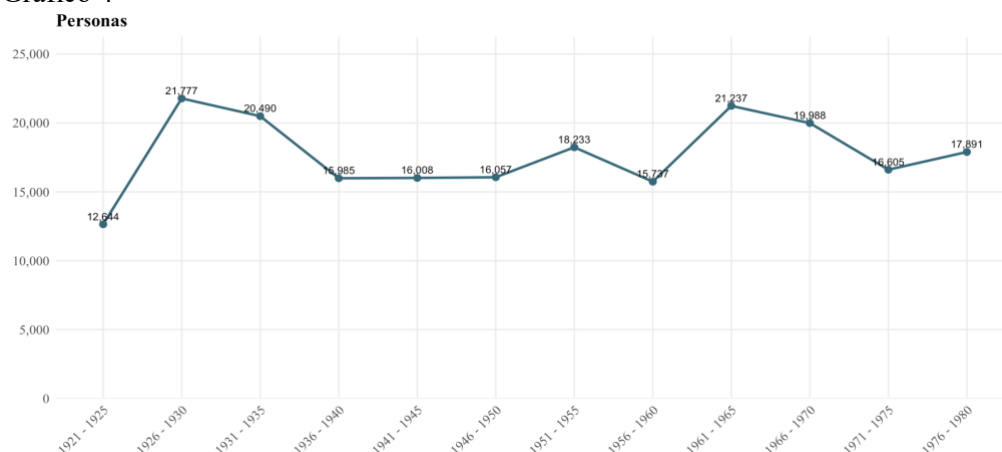
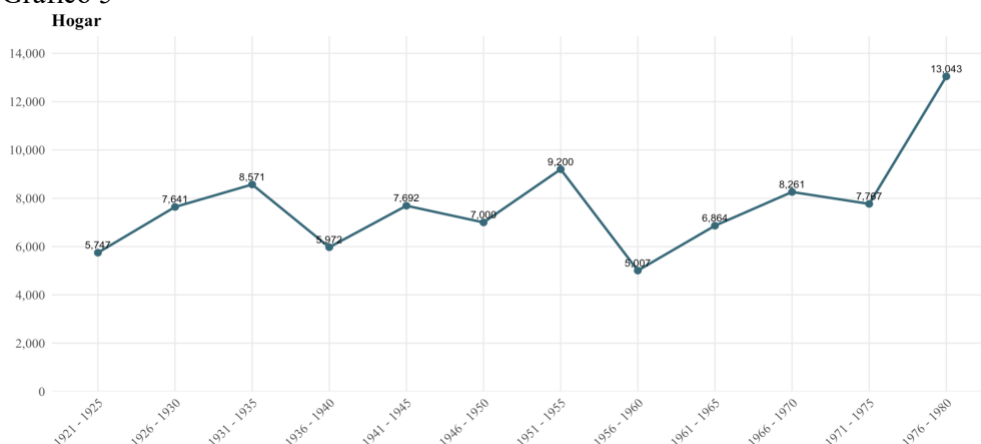


Gráfico 5



En la siguiente tabla (9) podemos ver los tópicos más significativos por letrista:

Letrista	Tópicos en que destaca
Alfredo Le Pera	Ansiedad, Muerte
Cátulo Castillo	Esparcimiento
Celedonio Flores	Dinero, Logro
Emilio Fresedo	Tristeza
Enrique Santos Discepolo	Religión
Francisco García Jiménez	Inhibición
Héctor Marcó	Cuerpo
Homero Expósito	Conflicto, Incertidumbre, Relaciones sociales
Horacio Sanguinetti	Familia
José González Castillo	Amigos, Entendimiento
Juan Andrés Caruso	Reflexión
Luis Rubinstein	Certeza, Emociones, Enojo
Pascual Contursi	Hogar, Personas, Salud, Sexualidad
Reinaldo Yiso	Trabajo

Los resultados de los tópicos LDA TF-IDF son los presentados a continuación, ordenados desde los más a los menos significativos, y organizados por generación.

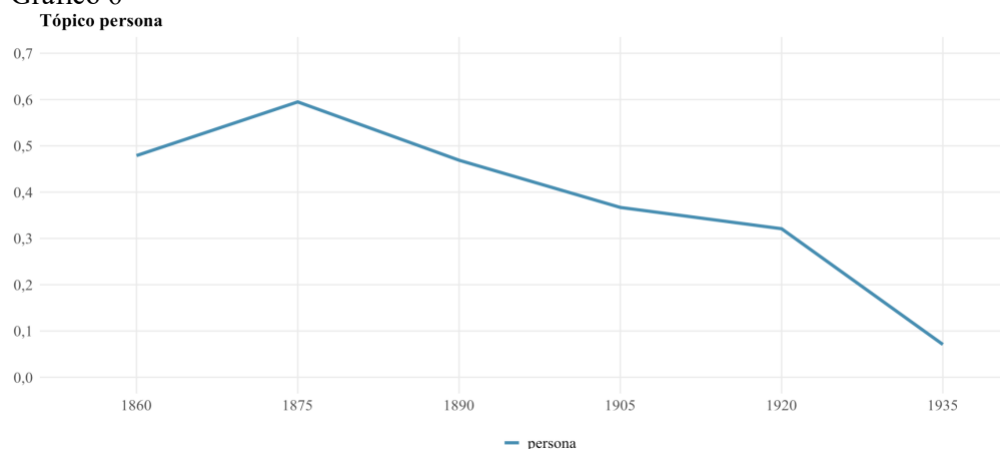
Tabla 10

Tópico	1860	1875	1890	1905	1920	1935	Promedio
Acción	0,493	0,449	0,432	0,459	0,410	0,571	0,469
Naturaleza	0,394	0,492	0,459	0,533	0,495	0,429	0,467
Lugar	0,324	0,552	0,487	0,525	0,458	0,429	0,462
Persona	0,479	0,595	0,469	0,367	0,321	0,071	0,384
Religión	0,507	0,423	0,358	0,411	0,382	0,143	0,371
Música	0,493	0,332	0,381	0,435	0,335	0,143	0,353
Tiempo	0,479	0,327	0,225	0,253	0,311	0,357	0,325
Objeto	0,211	0,254	0,282	0,251	0,321	0,143	0,244
Emoción	0,211	0,137	0,143	0,147	0,203	0,143	0,164
Flora	0,000	0,176	0,143	0,166	0,170	0,000	0,109
Fauna	0,000	0,100	0,108	0,104	0,118	0,071	0,084

Nota. Celdas muestran proporción por generación; el Promedio se recalcula como la media de 1860–1935.

No se observan en estos datos aspectos significativos que convenga considerar, con la sola excepción del descenso del tópico *persona*, tal como se presenta en Gráfico 6. Esto sin duda se debe al proceso de mayor abstracción del tango (Mina 2006) conforme transcurren las décadas, lo que produce una disminución de su carácter narrativo y, en consecuencia, un abandono de los personajes.

Gráfico 6



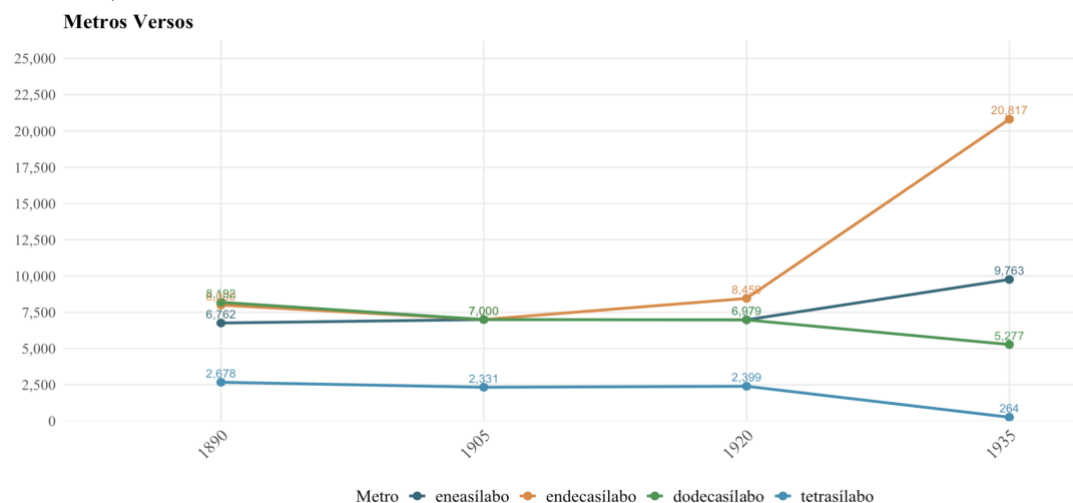
Respecto de las características métricas de los tangos, cuyas líricas o letras se asumen para el presente estudio como poemas, los hallazgos son los siguientes:

Tabla 11

Sílabas	Tipología	1890	1905	1920	1935	%
8	octosílabo	27,419	27,691	23,641	39,314	29,312
7	heptasílabo	12,735	12,360	14,849	7,124	12,369
6	hexasílabo	10,944	7,423	6,862	5,277	9,041
10	decasílabo	8,750	7,541	9,362	2,639	8,367
12	dodecasílabo	8,099	7,630	7,466	5,277	8,145
11	endecasílabo	6,989	7,893	8,456	20,317	7,074
9	eneasílabo	6,192	7,810	6,879	9,763	6,584
5	pentasílabo	6,246	6,228	7,181	2,639	6,085
13	tridecasílabo	2,809	3,337	3,003	2,902	2,861
14	tetradecasílabo	2,109	3,541	2,282	4,222	2,501
4	tetrasílabo	2,678	2,331	2,399	0,264	2,375
3	trisílabo	0,855	1,503	1,862	0,000	1,068
2	bisílabo	0,631	0,976	1,359	0,264	0,719

Se puede observar que los versos de arte menor –en particular los octosílabos, heptasílabos y hexasílabos– que abordan temas más ligeros –como ha mostrado la tradición analítica lírica desde Aristóteles (Martínez-Gamboa *et al.* 2025)– dominan los primeros lugares de los metros más ocupados –mientras que el endecasílabo, el metro canónico de arte mayor, solo ocupa un 7,1 % de los casos–, aglomerando más del 50% de los versos. Esto demuestra que el tango es, tal como mencionaba Borges, primero una aproximación liviana, jovial y festiva de la existencia, y solo más tarde aumenta su densidad temática, sobre todo desde la obra de letristas como Homero Manzi, Cátulo Castillo u Horacio Ferrer en tándem con Astor Piazzolla, llegando a su punto cúlmine en el electrotango de Bajofondo Tango Club (Romay 2000). Esto último se puede observar con más precisión en el siguiente gráfico.

Gráfico 7



Los tipos de estrofas se muestran en la siguiente tabla.

Tabla 12

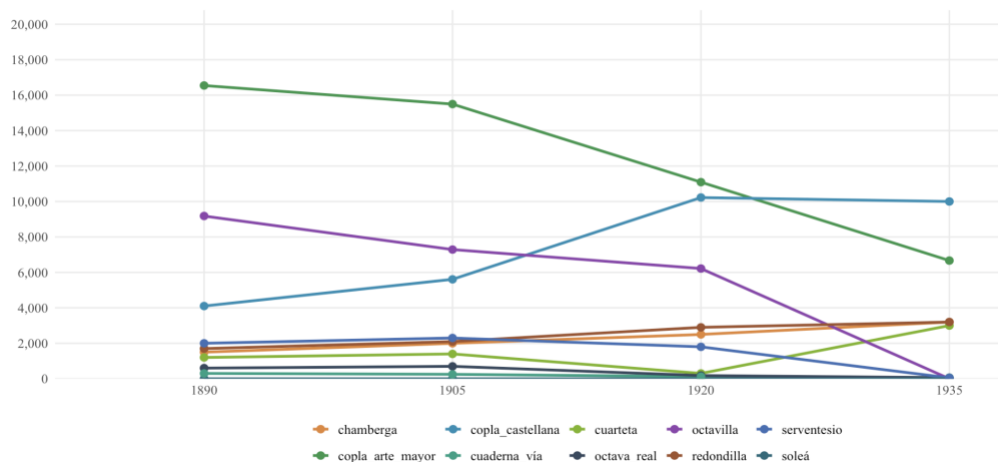
Distribución de estrofas por generación (con TOTAL)

ESTROFA	1890	1905	1920	1935	TOTAL
Octava	1.337	777	111	2	2.227
Copla arte mayor	1.218	835	66	2	2.121
Octavilla	676	401	37	0	1.114
Novena	498	452	33	3	986
Serventesio	302	278	61	3	644
Romance	357	232	23	0	612
Décima antigua	237	277	8	3	525
Quintilla	218	195	20	1	434
Sextilla	218	170	19	0	407
Couplet	185	165	18	0	368
Octava real	163	135	6	0	304
Seguidilla	176	108	18	1	303
Redondilla	120	144	18	1	283
Silva arromanzada	154	94	14	2	264
Sexteto	113	132	12	0	257
Cuarteta	139	98	1	0	238
Septeto	75	115	18	0	208
Quinteto	90	75	11	3	179
Seguidilla compuesta	76	64	19	0	159
Septilla	95	52	8	0	155

Dominando la octava a inicios del siglo XX, seguida por la copla de arte mayor, la octavilla y la novena, todos los tipos tradicionales de estrofas/stanzas van disminuyendo con el tiempo, ya que al son de los cambios epocales estos subgéneros se van abandonando para llegar al versolibrismo, que es propio de la época modernista (Godio y Mancuso 2006). Esto es un reflejo de la modernización de las ciudades y urbes, y el consiguiente abandono de las tradiciones rurales, coloniales y decimonónicas heredadas de la cultura española. En el caso de algunos metros, ello se puede graficar de la siguiente manera:

Gráfico 8

Evolución Estrofas



Esto, sin embargo, sucede en el tango, pero no en la milonga que, como ha explicado Jorge Drexler (2017) en una ilustrativa charla TED, sigue un patrón rítmico regular de ocho golpes (3 + 3 + 2) y tiende a ser octosílabo dactílico (esto es, dos pies dactílicos seguidos de un pie troqueo), como ilustran los dos primeros versos de la milonga rea “Sin novedad” (sin fecha precisa, Ubaldo Martínez): *Pase, agente Belisario, ¿hay alguna novedad? / Para decir la verdad, ninguna, mi comisario*. La tabla de estrofas de la milonga es la siguiente:

Tabla 13

ESTROFA	#	%
Octava	151	14,30
Octavilla	89	8,43
Décima_antigua	71	6,72
Unknown	66	6,25
Espinela	58	5,49
Quintilla	58	5,49
Cuarteta	52	4,92
Cantar	47	4,45
Novena	47	4,45
Redondilla	46	4,36
Romance	46	4,36
Serventesio	34	3,22
Seguidilla	28	2,65
Sextilla	28	2,65
Copla_arte_mayor	25	2,37
Sexteto	24	2,27
Couplet	18	1,70
Cuaderma vía	17	1,61
Quinteto	17	1,61
Copla_arte_menor	16	1,52
Otros	118	11,17

Junto con lo anterior, se puede observar cómo algunos tangueros específicos despliegan sus propias estrofas:

Tabla 14

Distribución de estrofas — Enrique
Cadícamo (1900)

Enrique Cadícamo (1900)	#	%
Copla arte mayor	152	20,026
Octava	97	12,780
Novena	85	11,199
Couplet	52	6,851
Quintilla	49	6,456
Octavilla	48	6,324
Redondilla	27	3,557
Décima antigua	25	3,294
Romance	25	3,294
Octava real	21	2,767
Sextilla	20	2,635
Copla arte menor	16	2,108
Copla castellana	13	1,713
Seguidilla compuesta	13	1,713
Serventesio	12	1,581
Quinteto	11	1,449
Sexteto	11	1,449
Cantar	8	1,054
Septeto	8	1,054
Cuaderna vía	6	0,791
Cuarteta	6	0,791
Soleá	6	0,791
Copla real	5	0,659
Cuarteto	5	0,659
Espinela	5	0,659
Endecha real	4	0,527
Estrofa Francisco de la Torre	4	0,527
Estrofa manriqueña	4	0,527
Ovillejo	4	0,527
Septilla	3	0,395
Terceto	3	0,395
Copla mixta	2	0,264
Haiku	2	0,264
Sexta rima	2	0,264
Chamberga	1	0,132
Estrofa Sáfica	1	0,132
Romance arte mayor	1	0,132
Seguidilla	1	0,132
Sexteto Lira	1	0,132
Total	759	100,000

Tabla 15

Homero Manzi (1907)	#	%
Unknown	93	12,97
Copla arte mayor	89	12,41
Octava	57	7,95
Octavilla	46	6,416
Serventesio	44	6,137
Romance arte mayor	39	5,439
Romance	34	4,742
Novena	29	4,045
Novena	29	4,045
Sextilla	26	3,626
Redondilla	24	3,347
Sexteto	22	3,068
Seguidilla	19	2,65
décima antigua	18	2,51
Octava real	18	2,51
Quintilla	16	2,232
Sexteto Lira	12	1,674
Cuarteta	11	1,534
Seguidilla compuesta	11	1,534
Couplet	10	1,395
Cuaderna vía	10	1,395
Silva arromanzada	10	1,395
Lira	10	1,395
Cantar	8	1,116
Sexta rima	8	1,116
Copla real	6	0,837
Haiku	5	0,697
Copla arte menor	2	0,279
Copla castellana	2	0,279
Cuarteto	2	0,279
Estrofa Francisco de la Torre	2	0,279
Chamberga	1	0,139
Endecha real	1	0,139
Quinteto	1	0,139
Septilla	1	0,139
Terceto monorrímo	1	0,139

Tabla 16

Celedonio Flores (1896)	#	%
Octavilla	54	10,36
Copla arte mayor	53	10,17
Serventesio	51	9,789
Octava	49	9,405
Quinteto	36	6,91
Unknown	30	5,758
Novena	21	4,031
Romance arte mayor	19	3,647
Couplet	18	3,455
Septilla	16	3,071
Seguidilla	16	3,071
Romance	15	2,879
Décima antigua	15	2,879
Sexteto	14	2,687
Redondilla	14	2,687
Cuarteta	13	2,495
Cantar	13	2,495
Ovillejo	12	2,303
Haiku	12	2,303
Chamberga	10	1,919
Copla castellana	9	1,727
Silva arromanzada	8	1,536
Cuaderna vía	7	1,344
Quintilla	6	1,152
Sextilla	5	0,96
Espinela	3	0,576
Terceto	2	0,384

Tabla 17

Enrique Santos Discépolo (1901)	#	%
Unknown	180	40,82
Octava	70	15,87
Romance	46	10,43
copla arte mayor	28	6,349
Novena	23	5,215
Sextilla	19	4,308
Décima antigua	15	3,401
Couplet	14	3,175
Septeto	11	2,494
Silva arromanzada	9	2,041
Ovillejo	8	1,814
Chamberga	7	1,587
Seguidilla	6	1,361
Quintilla	3	0,68
Cuaderna vía	1	0,227
Octavilla	1	0,227

Podemos determinar, según las tablas anteriores, que el letrista más versátil en cuanto a la variedad de estrofas/stanzas ocupadas es Enrique Cadícamo, mientras que el mayor versolibrista de los cuatro es Enrique Santos Discépolo. Finalmente, los tópicos predilectos para la escritura de algunos tipos de estrofa/stanzas son los que se muestran en Tabla 18 a continuación:

TIPO DE ESTROFA	TÓPICO EN QUE DESTACA
Cuarteta	Enojo
Octava real	Reflexión
Octavilla	Relaciones sociales
Quinteto	Esparcimiento
Redondilla	Incertidumbre
Romance	Hogar
Romance arte mayor	Logro
Seguidilla	Personas
Seguidilla compuesta	Emociones
Septeto	Cuerpo
Serventesio	Familia
Sexteto	Conflicto
Silva arromanzada	Entendimiento

El proceso de detección de estos arrojó la presencia de 227 lunfardimos puros diferentes en el corpus. La siguiente presenta los más frecuentes en la colección:

Tabla 19

LUNFARDISMO	CASOS
pa	1084
malevo	325
vento	159
taita	124
gotán	119
metejón	102
catrera	98
chamuyo	94
compadrino	91
pilcha	88
bailongo	87
taura	85
paica	81
purrete	72
fueir	70
engrupir	68
rante	66
tapera	63
farolito	57
ma	56

A su vez, la tabla siguiente presenta a los veinticuatro letristas con más letras en el corpus, junto con el porcentaje de uso del lunfardo en sus obras. Se calculó el porcentaje considerando el total de tangos de cada letrista que presentan al menos un lunfardismo, dividido por el total de tangos escritos por este:

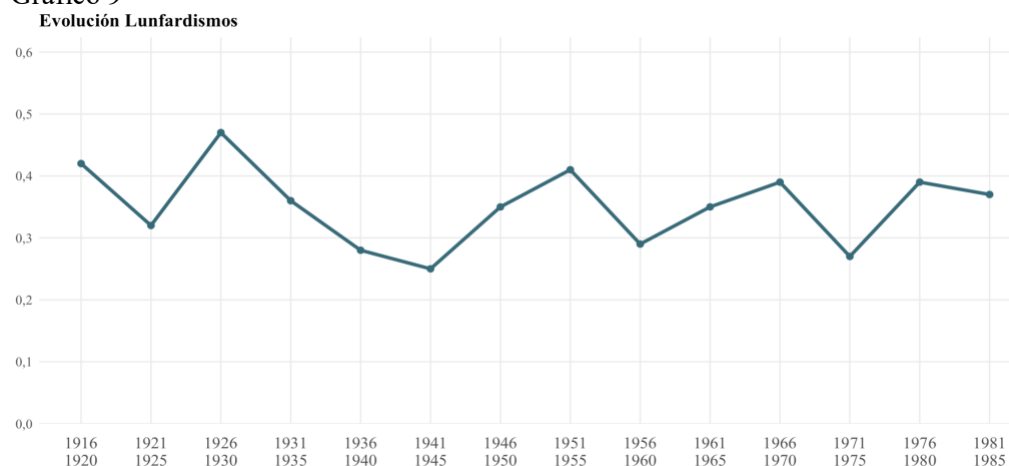
Tabla 20

Letristas: cantidad de tangos y proporción de lunfardismos

LETRISTA	#TANGOS	% LUNFARDISMOS
Pascual Contursi	83	0,99
Celedonio Flores	153	0,67
Manuel Romero	117	0,56
Cátulo Castillo	101	0,51
Gabriel Clausi	59	0,49
Enrique Santos Discepolo	94	0,49
Juan Andrés Caruso	120	0,48
Alfredo Le Pera	70	0,44
Emilio Fresedo	76	0,41
Enrique Cadícamo	289	0,39
Lito Bayardo	57	0,35
Homero Manzi	201	0,32
Héctor Marcó	85	0,32
Enrique Dizeo	57	0,32
Francisco García Jiménez	154	0,31
Ivo Pelay	78	0,29
Homero Expósito	105	0,28
Luis Rubinstein	98	0,26
José González Castillo	72	0,25
Reinaldo Yiso	77	0,25
Horacio Sanguinetti	82	0,24
José Rótulo	57	0,18
Carlos Bahr	167	0,17
José María Contursi	146	0,14

Es de hacer notar que el letrista más lunfardero sea justamente Pascual Contursi, el creador del tango-canción (Mina 2006) y modelo a seguir por todos los letristas posteriores. Finalmente, la evolución de los tangos con lunfardismos por periodos es la siguiente:

Gráfico 9



Podemos observar que la presencia de lunfardismos experimenta oscilaciones en el tiempo, pero que, desde una mirada más detallada, alcanza su momento más significativo en el periodo 1926-1930, el de la Guardia Nueva capitaneada por Carlos Gardel; y su momento de menos significancia en el periodo inicial de las Grandes Orquestas. En efecto, entre 1926 y 1945 se observa un sostenido descenso en el uso de lunfardismos en las letras, asociado, como ha notado Mina (2006), al cambio epocal que abarca desde el tango arrabalero al momento de la nostalgia del paraíso perdido, donde hay más un recuerdo del barrio, pero sin el argot propio de la cultura del Río de la Plata de inicios del siglo XX. Esto también se debe, como también señala Mina (2006), a la prohibición del lunfardo en el periodo desde 1943 hasta 1949.

Por otro lado, el ascenso y descenso oscilante de los periodos posteriores a 1945 bien puede deberse a una espectacularización (Debord 2003) en la que el uso del lunfardo ya no es un uso cotidiano, sino que un anacronismo que tiñe a los tangos posteriores de una fantasía arrabalera que no se condice con las circunstancias auténticas de su producción.

Conclusiones

Las conclusiones de este estudio interdisciplinario sobre las letras del tango emergen de la integración sistemática de hallazgos métricos, emocionales y temáticos analizados a través de un corpus de 6.811 obras. Lejos de ser una mera síntesis, el análisis computacional avanzado (utilizando SICAM, el modelo NES para polaridad y LIWC/LDA para tópicos) permite establecer correlaciones profundas entre la evolución formal del género y su capacidad para reflejar las pasiones y tensiones de la modernidad urbana rioplatense.

Uno de los aportes más significativos del estudio es la confirmación cuantitativa de la tesis de la melancolía del tango. Los resultados de polaridad muestran una tendencia clara y sostenida hacia una mayor negatividad emocional en las letras a medida que avanza el tiempo. Esta evolución refrenda la observación cultural de Jorge Luis Borges (2016) de que el tango, inicialmente un baile “valeroso y feliz” derivado de la milonga, progresivamente “va languideciendo y entristeciéndose”, lectura que es confirmada por Castro (2007) y Schvartzman (2020). Este cambio afectivo no ocurre de forma aislada, sino que se encuentra íntimamente ligado a la transformación estilística y métrica del género.

En sus primeras etapas, el tango se caracteriza por el dominio de los versos de arte menor, especialmente el octosílabo, el heptasílabo y el hexasílabo, los que, en conjunto, aglomeran más del 50% de los versos. Esta preponderancia inicial del arte menor, tradicionalmente asociado en la lírica con temas más ligeros y joviales, se corresponde perfectamente con el carácter festivo inicial del género. Sin embargo, el estudio demuestra que, con el paso de las décadas, existe una disminución de los tipos tradicionales de estrofas y metros de arte menor, a la vez que las estructuras más complejas y el versolibrismo ganan terreno. Este movimiento hacia formas métricas más libres y sofisticadas, como el versolibrismo de letristas como Enrique Santos Discépolo, es un reflejo poético de la modernización de las ciudades del Río de la Plata y del abandono de las tradiciones rurales, coloniales y decimonónicas.

La evolución emocional y formal se ve espejada en los cambios temáticos identificados por LIWC y LDA TF-IDF. El análisis de tópicos revela un aumento en la mención de *emociones*, *muerte* y *trabajo*. El incremento del tópico *muerte* se interpreta como una probable metáfora asociada a la pérdida del “Buenos Aires querido”, relacionado con el tema lírico del paraíso perdido que no solo se observa en tangos tardíos como “Sur” (1948,

Homero Manzi y Aníbal Troilo) o “Tinta roja” (1941, Catulo Castillo y Sebastián Piana), sino que en otros más tempranos como “Corrientes y Esmeralda” (1933, Celedonio Flores y Francisco Pracánico), y la añoranza, un sentimiento reforzado por el aumento general de la categoría *emociones*. Paralelamente, el ascenso del tópico *trabajo* se vincula al progreso de las sociedades rioplatenses y a la disminución de los oficios esporádicos en favor de estructuras laborales más formales. En contraste, la disminución del tópico *salud* sugiere que el desarrollo de la medicina y la eficiencia de los servicios hospitalarios en la modernidad urbana hicieron que las enfermedades fueran menos relevantes en la vida diaria, sacándolas del foco lírico.

Desde la perspectiva del modelado LDA TF-IDF, se observa una disminución significativa del tópico *persona* con el tiempo. Este hallazgo subraya un proceso de mayor abstracción en el tango (Mina 2006), en que el género abandona progresivamente su carácter narrativo y la centralidad de personajes concretos.

El análisis del lunfardo ofrece una perspectiva matizada sobre la autenticidad y la representación del arrabal. Aunque Pascual Contursi, considerado el creador del tango canción, es el letrista más “lunfardero” del corpus, el uso de lunfardismos puros alcanza su momento más significativo durante el periodo 1926-1930 (la Nueva Guardia, asociada a Gardel). Sin embargo, posteriormente observamos un descenso sostenido hasta 1945. La oscilación posterior a 1945 sugiere un fenómeno de espectacularización en el que el lunfardo deja de ser un uso cotidiano y se convierte en un anacronismo que tiñe los tangos de una “fantasía arrabalera” que no necesariamente se corresponde con las circunstancias contemporáneas de su producción.

Lo que se puede observar y destacar como hallazgo principal y general del estudio es que efectivamente, y tal como otros géneros del inicio de la canción popular en Occidente — como la canción napolitana italiana, la canción de Montmartre francesa, el Tin Pan Alley estadounidense, el *music hall* británico o el cuplé español —, el tango contribuye a construir las identidades, las mentalidades y las formas de vida de un pueblo —en este caso, el rioplatense sobre todo de la primera mitad del siglo XX—, así como sus transformaciones y, en ese sentido, resulta una ventana a la historia social y cultural de su lugar y su tiempo.

Finalmente, el uso de la metodología computacional permite establecer perfiles autorales detallados. Por ejemplo, es posible determinar la versatilidad métrica de autores como Enrique Cadícamo en contraste con la inclinación al versolibrismo de Enrique Santos Discépolo, y cómo letristas específicos, como Pascual Contursi u Homero Manzi, se inclinan por tópicos predilectos. Este enfoque va más allá del análisis estilístico cualitativo y ofrece una visión cuantificada del modo en que las generaciones de letristas se alinearon con los cambios formales y temáticos del género.

En definitiva, esta investigación contribuye al estudio de la lírica tanguera al proveer una cartografía sistemática de la relación entre métrica, afectos y evolución temática, lo que demuestra que las transformaciones estéticas del tango son un espejo cuantificable de las complejidades sociales y emocionales de la modernidad urbana en el Río de la Plata.

Con todo, los hallazgos no resultan disímiles de lo que han concluido quienes han llevado a cabo estudios analíticos y críticos cualitativos respecto del género (Romano 1991; Ulla 1982; Vilariño 1981, 1995). Debe, sin embargo, enfatizarse que este tipo de metodologías corresponde a data robusta en los análisis que lleva a cabo, por lo que, al refrendar lo que han logrado los estudios cualitativos, aporta un respaldo significativo a los mismos, aunque, en ocasiones, puede potencialmente rechazar dichas tesis, lo que no resulta ser siempre el caso en este mismo estudio. Por ejemplo, el modelado de tópicos ha encontrado

que la religión resulta un tema propio de las letras de Enrique Santos Discépolo, aunque esto no se condice con la investigación cualitativa. Debe, entonces, precisarse que, en los casos en que los hallazgos cualitativos sean contradictorios con los cuantitativos, estos deben someterse a un riguroso análisis cualitativo posterior, tomando los datos en contradicción, para arribar a respuestas más seguras. Esperamos que este análisis abra puertas a nuevas vías para el estudio de la canción popular, y permita redescubrir su riqueza y su pertinencia para comprender la vida moderna.

Bibliografía

- Atkinson-Abutridy, Jaime. 2022. *Text Analytics: An Introduction to the Science and Applications of Unstructured Information Analysis*. Nueva York: Chapman and Hall/CRC.
- Borges, Jorge Luis. 2016. *El tango. Cuatro conferencias*. Buenos Aires: Sudamericana.
- Castro, Fernando Andrés. 2007. "Tango: melancolía que se baila en nuestro tiempo del desprecio". *Folios, Revista de la Facultad de Comunicaciones y Filología*, (12-13), 46-51.
- Conde, Óscar. 2011. *Diccionario etimológico del lunfardo*. Buenos Aires: Taurus.
- Coupland, Douglas. 2001. "Generation X" en *Postmodern American Fiction: A Norton Anthology*, editado por Paula Geyh, Fred G. Leebron y Andrew Levy. Nueva York: Norton: 568-573.
- De Mayo, Oscar. 1998. "La evolución del tango y la identidad ciudadana". *Signos Universitarios*, 18 (34).
- Debord, Guy. 2003 [1967]. *La sociedad del espectáculo*. Valencia: Pre-Textos.
- Drexler, Jorge. 2017. "Poetry, Music and Identity". *TED*.
- Espíndola, Alejandro. 2002. *Diccionario del lunfardo*. Buenos Aires: Planeta.
- García, Ricardo. 2008a. "La cumparsita". *Todo Tango* (web).
- _____. 2008b. "Reflexiones sobre los orígenes del tango". *Todo Tango* (web).
- Gobello, José. 1967. "El difícil nacimiento del tango". *Revista Sucesos. Serie Hechos Mundiales*, 4 (septiembre). Santiago: Zig-Zag.
- Gobello, José. 1991. *Nuevo diccionario lunfardo*. Buenos Aires: Corregidor.
- Godio, Julio y Héctor Mancuso. 2006. *La anomalía argentina. De la tierra prometida a los laberintos de la frustración*. Buenos Aires: Miño y Dávila.
- González, Juan Pablo y Claudio Rolle. 2004. *Historia social de la música popular chilena, 1890-1950*. Santiago de Chile: Ediciones Universidad Católica de Chile.
- Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li y Liang Zhao. 2019. "Asignación de Dirichlet Latente (LDA) y modelado de tópicos: modelos, aplicaciones, una encuesta". *Multimedia Tools and Applications*, 78: 15.169-15.211. <https://doi.org/10.1007/s11042-018-6894-4>
- Lage, Lola. 2014. "Herramienta para el análisis de la opinión en tweets periodísticos". Trabajo de grado en Ingeniería Informática, Universitat Pompeu Fabra, Barcelona. https://repositori.upf.edu/bitstream/handle/10230/22939/LageGarcia_2014.pdf?sequence=1&isAllowed=y
- Liska, María. 2008. "La renovación de los valores de autenticidad o el tango electrónico como producto del mercado". Ponencia presentada en el VIII Congreso IASPM-AL, Lima.

- Martínez-Gamboa, Ricardo. 2019. *Clásicos AM: historia de la balada romántica latinoamericana*. Santiago de Chile: Planeta.
- Martínez-Gamboa, Ricardo, Suyli Aproso, Valentina Araya-Lopetegui y Nicolás Acosta. 2025. "Análisis de las correlaciones entre métrica y polaridad emocional en un corpus de poemas chilenos (siglos XIX-XXI): una aproximación desde el análisis literario computacional". *Revista la Palabra*, 50.
- Mina, Carlos. 2006. *Tango. La mezcla milagrosa (1917-1956)*. Buenos Aires: Sudamericana.
- Monjeau, Federico. 1998. "Cada día cantan mejor". *Clarín*, 1º de julio de 1998.
- Morales Pettorino, Felipe. 1984. *Diccionario ejemplificado de chilenismos*. Valparaíso: Academia Superior de Ciencias Pedagógicas de Valparaíso.
- Navarro, Jorge. 2008. "La historia del tango". *Todo Tango* (web).
- Nikolenko, Sergey I., Sergei Koltcov y Olga Koltsova. 2017. "Topic Modelling for Qualitative Studies". *Journal of Information Science*, 43 (1): 88-102.
- Puppo, María Eugenia, Carolina Nóbile y Ana Sader. 2022. "Urbanismo epidémico. La medicalización de la vida urbana en Montevideo en el primer tercio del siglo XX". *Ciudades*, 25: 153-179.
- Ramírez-Esparza, Nairán, James W. Pennebaker, Fernando A. García y Rocío Suría. 2007. "La psicología del uso de las palabras: un programa de computadora que analiza textos en español". *Revista Mexicana de Psicología*, 24 (1): 85-99.
- Real Academia Española. s. f. *Diccionario de la lengua española*, 23ª ed. [versión en línea]. <https://dle.rae.es>
- Romano, Eduardo. (Ed.). (1991). *Las letras del tango: antología cronológica 1900-1980*. Rosario: Editorial Fundación Ross.
- Romay, Horacio. 2000. *El tango y sus protagonistas*. Buenos Aires: Bureau.
- Rosati, Guido. 2022. "Procesamiento de lenguaje natural aplicado a las ciencias sociales: detección de tópicos en letras de tango". *Revista Latinoamericana de Metodología de la Investigación Social*, 23 (12): 38-60.
- Sarlo, Beatriz. 1995. *Borges, un escritor en las orillas*. Buenos Aires: Ariel.
- Schvartzman, Julio. 2020. "Borges y su tango triste". *Chuy. Revista de Estudios Literarios Latinoamericanos*, 7(9), 206-242.
- Ulla, Noemí. (1982). *Tango, rebelión y nostalgia*. Centro Editor de América Latina.
- Vilariño, Idea. (1981). *El tango cantado*. Buenos Aires: Calicanto Editorial.
- _____. (1995). *El tango: estudio y antología*. Montevideo: Cal y Canto.