

NLP Informe: Taller de la Ley de Zipf

Aponte Duque, Juan
aponte.juan@javeriana.edu.co

Cardozo Amin, Cristian
cristiancardozo@javeriana.edu.co

Triana Ocampo, Benkos
b.triana@javeriana.edu.co

10 de agosto de 2025

Resumen

La Ley de Zipf es un patrón estadístico que aparece en cualquier idioma. De forma sencilla, establece que la segunda palabra más frecuente en un texto aparecerá aproximadamente con la mitad de la frecuencia de la primera; la tercera, con un tercio; y así sucesivamente, hasta que la palabra n -ésima aparezca con una frecuencia cercana a $1/n$ de la más frecuente. (Wikipedia, 2025g)

En este trabajo buscamos comprobar esta ley en un conjunto de canciones en español. Además, analizamos cómo varían los resultados al incorporar aspectos de fonética y fonología al análisis, evaluando si la relación propuesta por la Ley de Zipf se mantiene o se modifica, y observando el comportamiento de la distribución de las palabras bajo estas condiciones adicionales.

1. Introducción

La **Ley de Zipf** establece que, si ordenamos las palabras de un texto de mayor a menor frecuencia, la segunda palabra más usada aparecerá aproximadamente la mitad de veces que la primera, la tercera un tercio, y así sucesivamente. Este patrón se observa en distintos tipos de textos, como libros, discursos y, por supuesto, canciones.

Las letras de canciones son un caso interesante para analizar esta ley, ya que suelen contener repeticiones frecuentes: coros, frases cortas, e incluso sonidos no

léxicos como “ay” u “oh”. Estas características las convierten en un buen material para observar y poner a prueba la relación propuesta por Zipf.

Objetivos del taller:

1. Probar diferentes métodos de segmentación de palabras.
2. Contar las palabras y calcular su frecuencia de aparición.

2. Datos y fuentes

Para este taller, se recogió una muestra de canciones populares en Colombia, seleccionando una por cada año, desde 1945 hasta 2025, siempre que fuera posible. La elección priorizó la canción más reconocida o significativa del año. En los casos en que la letra no estaba disponible, se omitió ese año.

Toda la información se organizó en una excel, con columnas para *año*, *título*, *artista*, *género* y *letra*. Para ello, se consultaron tres fuentes principales:

- **Las 50 mejores canciones de Colombia**, lista publicada por el diario *El Tiempo* (El Tiempo, 2012).
- **Radio Nacional de Colombia: 80 años en 80 canciones**, recopilación oficial de la emisora (Radio Nacional de Colombia, 2014).
- El video **FULANITOVIAJERO Periodista Musical**, que analiza la historia musical del país (FULANITOVIAJERO Periodista Musical, 2023).

El resultado es un listado diverso que abarca múltiples géneros porro, cumbia, bambuco, salsa, vallenato, balada, rock, pop, reggeton, entre otros y refleja tanto la evolución de la música colombiana como las tendencias de cada época. Esta variedad permite analizar si la Ley de Zipf se cumple en estilos musicales distintos y cómo la repetición de ciertos géneros puede influir en la distribución de frecuencias de las palabras.

La lista completa de canciones seleccionadas, organizada por año, género, título y artista, se incluye en el **Apéndice A**.

2.1. Distribución de géneros

Para entender mejor la muestra, calculamos la distribución de géneros musical. La Figura 1 resume la proporción de cada género dentro del conjunto.

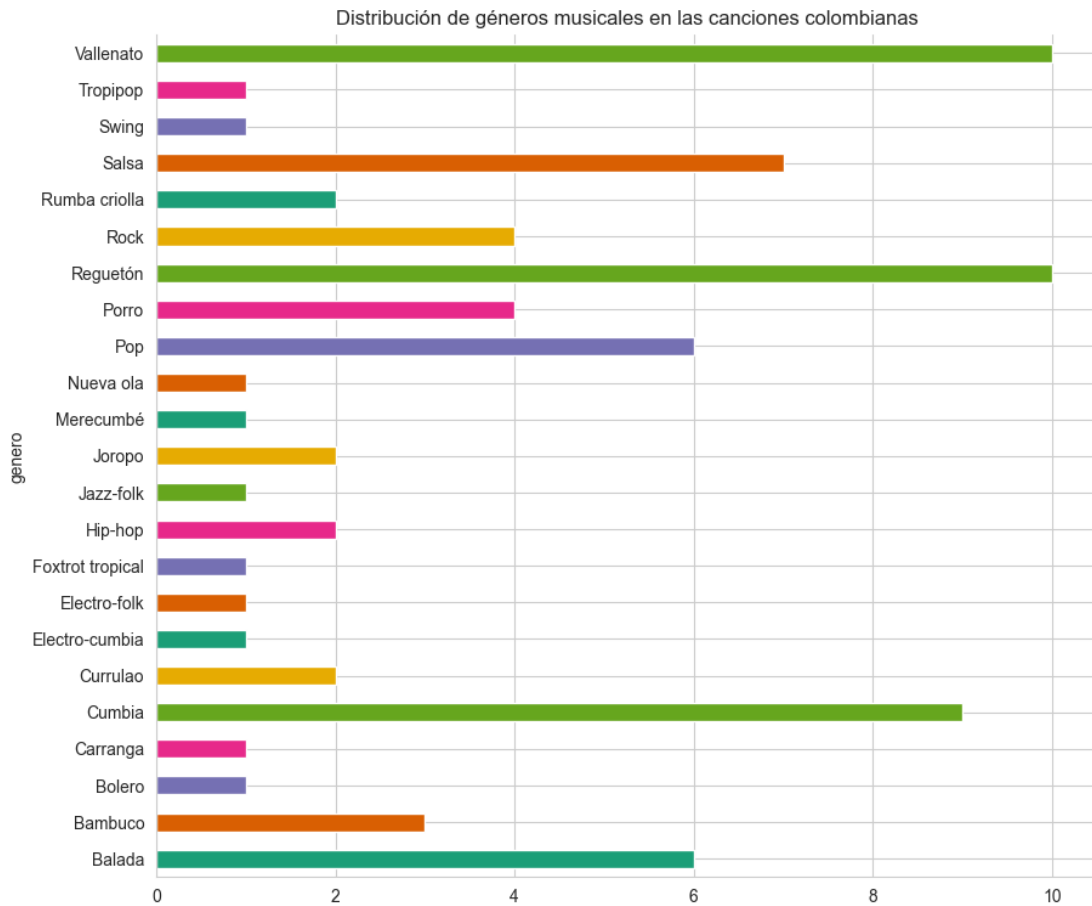


Figura 1: Distribución de géneros.

3. Análisis léxico y validación de la Ley de Zipf

En esta etapa, el objetivo principal es comprobar si las letras de las canciones cumplen el patrón descrito por la Ley de Zipf. Para ello, se realiza un análisis léxico que incluye:

1. **Tokenización:** dividir el texto de cada canción en palabras o tokens individuales.
2. **Conteo de frecuencias:** calcular cuántas veces aparece cada token en las canciones.
3. **Visualización:** graficar la relación entre la posición en el ranking y la frecuencia de aparición.

3.1. Opciones de tokenización evaluadas

Para la división del texto en tokens, se probaron distintas herramientas y enfoques:

- **Spanish + NLTK:** tokenizador en español incluido en NLTK.
- **TweetTokenizer:** diseñado para textos cortos y redes sociales, capaz de manejar emoticonos y símbolos especiales.
- **BERT (base):** tokenizador de un modelo de lenguaje general entrenado en múltiples idiomas.
- **RoBERTa entrenado en corpus en español:** versión optimizada para el Español.

Tras las pruebas, se eligió el tokenizador en español de NLTK. La decisión se basó en que no modifica los tokens originales (lo que conserva la forma exacta de las palabras) y facilita la interpretación de los resultados.

3.2. Ejemplo comparativo de tokenización

Para ilustrar las diferencias, en la Tabla 1 se muestra un fragmento de la canción *Colombia Tierra Querida* tokenizado con dos métodos: BERT y NLTK.

BERT	NLTK (español)
[Colombia, ,, tier, ##ra, que, ##rida, ,, him, ##no, de]	[Colombia, ,, tierra, querida, ,, himno, de, fe, y, armonía]
[f, ##e, y, arm, ##on, ##ía, ., Can, ##tem, ##os]	[., Cantemos, ,, cantemos, todos, grito, de, paz, y, alegría]
[,, can, ##tem, ##os, to, ##dos, g, ##rito, de, p]	[., Vivemos, ,, siempre, vivimos, a, nuestra, patria, querida, .]

Cuadro 1: Comparación de tokenización de un fragmento con BERT y NLTK.

Como se observa, BERT tokeniza agregando el prefijo **##** y dividiéndola en 2 la palabra, lo que genera una mayor fragmentación y requiere pasos adicionales para un análisis léxico. Por ejemplo, *tierra* se divide en *tier* y *##ra*, y *querida* en *que* y *##rida*.

En contraste, el tokenizador de NLTK conserva las palabras completas.

4. Frecuencia de palabras

En la Figura 2 se muestra la frecuencia relativa de las 100 palabras más comunes. Se observa un patrón claro: unas pocas palabras concentran gran parte de las ocurrencias, mientras que la mayoría tienen frecuencias mucho menores.

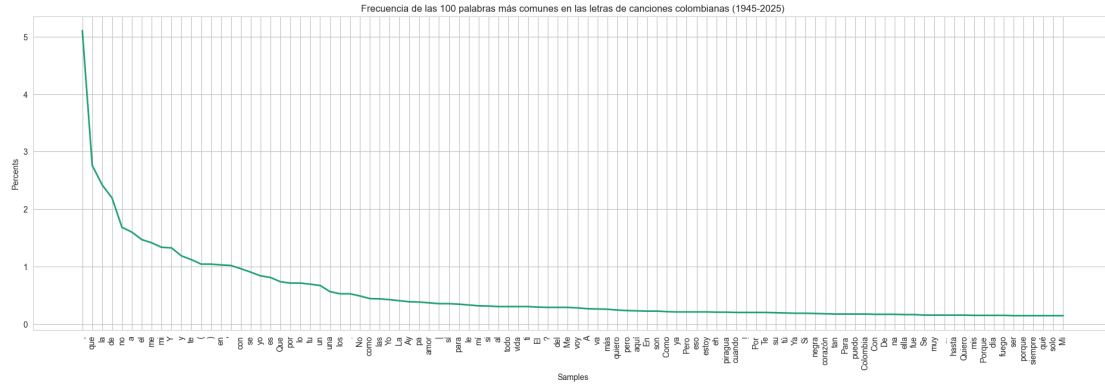


Figura 2: Frecuencia relativa (%) de las 100 palabras más comunes en el corpus.

La Figura 3 presenta la frecuencia acumulada.

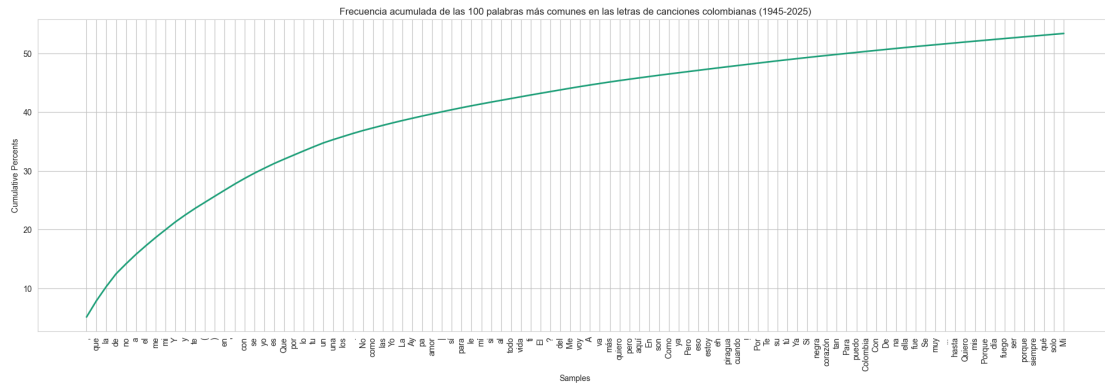


Figura 3: Frecuencia acumulada (%) de las 100 palabras más comunes en el corpus.

Este comportamiento es consistente con lo que predice la Ley de Zipf: un pequeño grupo de palabras domina el texto, mientras que la mayoría aparece de forma esporádica.

5. Validación de la Ley de Zipf

La Figura 4 presenta la relación entre el ranking y su frecuencia.

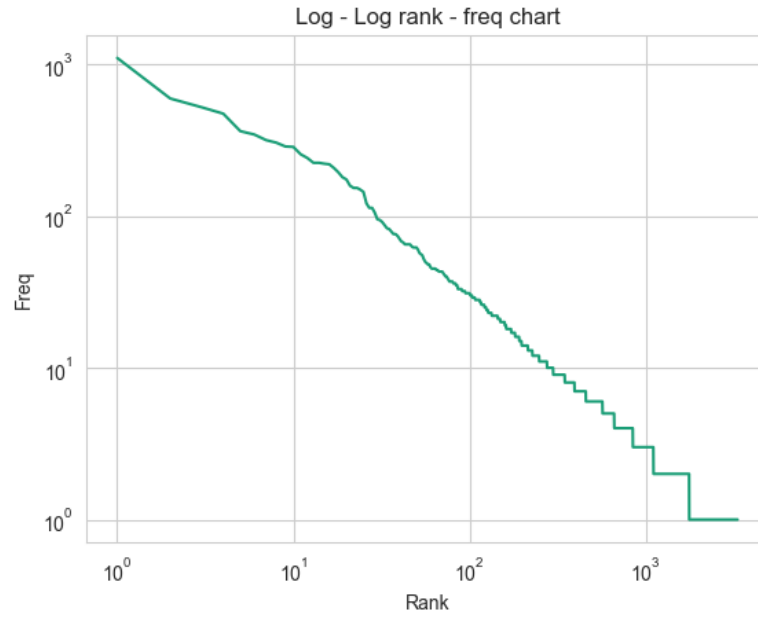


Figura 4: Ranking y Frecuencia en escala log-log.

Dado que, las canciones siguen la Ley de Zipf, los puntos deberían alinearse aproximadamente formando una línea recta descendente en la escala log-log.

5.1. Frecuencia relativa sin *stop words*

Se repitió el cálculo de frecuencias excluyendo las *stop words*, es decir, aquellas palabras muy comunes y sin carga semantica (por ejemplo: *el, la, de, y*).

La Figura 5 muestra la frecuencia relativa de las 100 palabras más usadas después de aplicar esta limpieza.

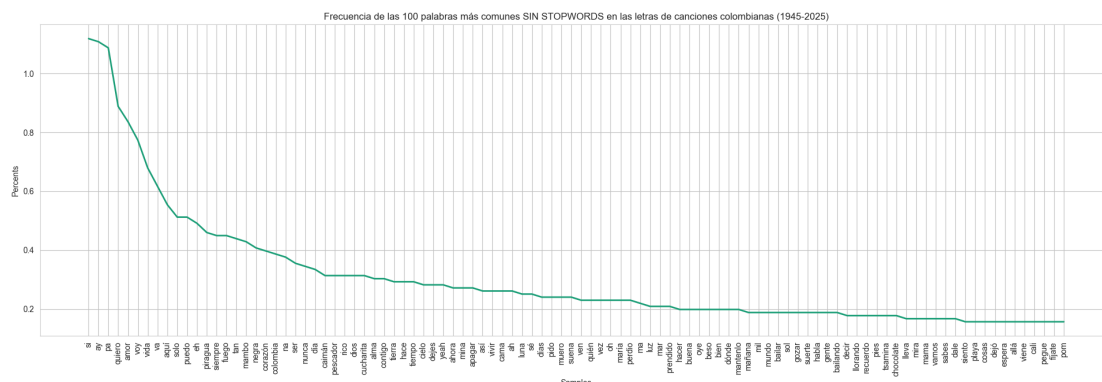


Figura 5: Frecuencia relativa de las 100 palabras más comunes excluyendo *stop words*.

Tras eliminar las *stop words*, la lista de palabras más frecuentes refleja mejor el contenido de las canciones, donde las palabras quiero, amor y vida son muy frecuentes.

6. Análisis fonológico y fonético

Ahora bien, queremos ver las características sonoras de las letras. Para ello, cada palabra del conjunto de canciones fue transcrita fonéticamente utilizando el Alfabeto Fonético Internacional.

Esto nos permite identificar patrones que no siempre son visibles en la escritura estándar. Mientras que el análisis léxico se centra en las unidades gráficas (palabras escritas), el análisis fonológico y fonético se centra en fonemas y sílabas. Este proceso lo vamos a realizar en varios pasos

1. **Ejemplo de transcripción fonológica y fonética:** Utilizaremos una palabra de ejemplo para ver como es el funcionamiento de la librería **fonemas**.
2. **Conversión fonética:** uso de herramientas automáticas de transcripción a IPA para el español.
3. **Extracción de patrones:** identificación de repeticiones fonéticas, estructuras silábicas predominantes y posibles rimas internas.

6.1. Ejemplo de transcripción fonológica y fonética

Se utilizó como palabra de prueba *esternocleidomastoideo*, un término largo y fonéticamente complejo. El resultado obtenido fue:

- **Transcripción fonológica (palabra completa):** ['esrnoklejdomastoj'deo']
- **Transcripción fonética (palabra completa):** ['esrnoklejðomastoj'ðeo']
- **Transcripción fonológica (sílabas):** ['es', 'ter', 'no', 'klej', 'do', 'mas', 'toj', 'de', 'o']
- **Transcripción fonética (sílabas):** ['es', 'ter', 'no', 'klej', 'ðo', 'mas', 'toj', 'ðe', 'o']

Este ejemplo permite visualizar:

1. **Diferencias entre fonología y fonética:** La fonológica representa los sonidos básicos (fonemas) de la palabra, mientras que la fonética refleja la pronunciación real, incluyendo variaciones como el paso de [d] a [ð].
2. **Estructura silábica:** En las sílabas nos permite analizar la distribución de patrones sonoros, detectar secuencias frecuentes.

7. Frecuencia de fonemas y sílabas

Ahora bien, antes se realizó un proceso de limpieza para eliminar signos de puntuación y otros símbolos no lingüísticos (por ejemplo: ., ,, !, ?). Además, se quitaron removieron los *stop words*.

Con las transcripciones fonológicas y fonéticas obtenidas para todas las canciones, se calcularon las frecuencias relativas de los elementos básicos del habla: fonemas y sílabas.

El análisis se realizó de manera separada para:

1. Fonemas fonológicos.
2. Fonemas fonéticos.
3. Sílabas fonológicas.
4. Sílabas fonéticas.

7.1. Fonemas fonológicos

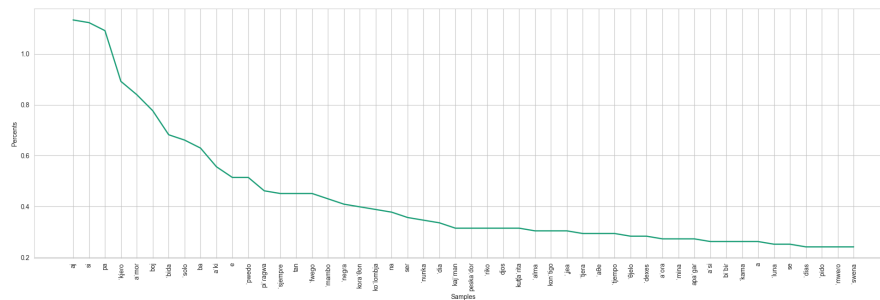


Figura 6: Frecuencia relativa de fonemas fonológicos.

7.2. Fonemas fonéticos

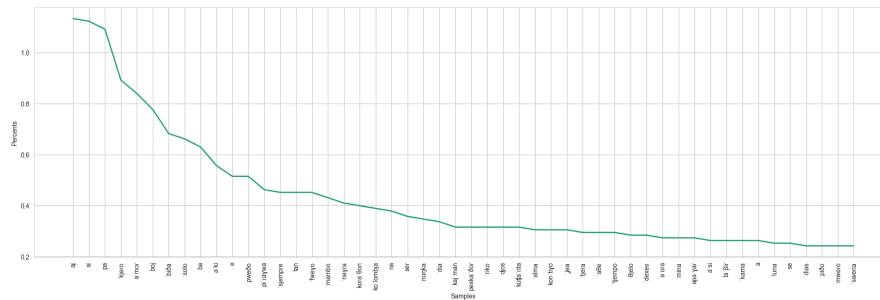


Figura 7: Frecuencia relativa de fonemas fonéticos.

7.3. Sílabas fonológicas

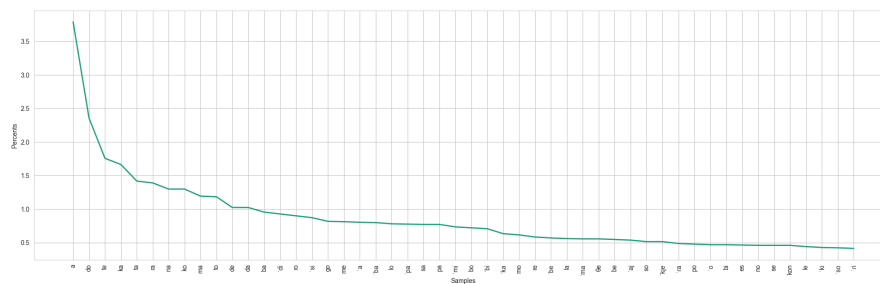


Figura 8: Frecuencia relativa de sílabas fonológicas.

7.4. Sílabas fonéticas

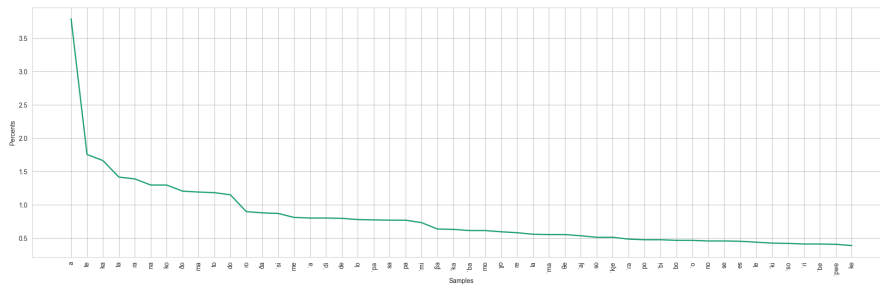


Figura 9: Frecuencia relativa de sílabas fonéticas.

Se observa que, igual que con las palabras, unas pocas unidades sonoras concentran la mayor parte de las apariciones. Esto confirma un patrón tipo Zipf también a nivel fonológico y fonético: pocos sonidos muy usados y muchos sonidos poco frecuentes.

Además, ciertos fonemas aparecen asociados a varias palabras diferentes, muchas de ellas expresiones cortas típicas en canciones. Por ejemplo, el fonema *be* se encuentra en *vé*, *we*, *be* y *ve*; mientras que *u* aparece como *uh*, *uhh* y *huh*.

A modo de ejemplo, la siguiente tabla resume algunos de los fonemas más recurrentes y ejemplos de palabras asociadas:

fonología	palabras asociadas	numero de palabras
be	<i>vé, we, be, ve</i>	4
u	<i>uh, uhh, huh</i>	3
aj	<i>ai, ay</i>	2
ej	<i>hey, ey</i>	2
'solo	<i>sólo, solo</i>	2
te	<i>the, t</i>	2
tan	<i>tan, than</i>	2
ma	<i>má, ma</i>	2
is	<i>is, his</i>	2
e	<i>eh, ehh</i>	2
ba	<i>wa, va</i>	2
de	<i>d, dé</i>	2
bjen	<i>bién, bien</i>	2
bi	<i>ví, vi</i>	2
bas	<i>vas, was</i>	2
ban	<i>van, wan</i>	2
'jɛa (aprox.)	<i>yeah, yeahhh</i>	2

Cuadro 2: Ejemplos de fonemas frecuentes y palabras asociadas.

8. Análisis Textométrico

La **textometría** es el conjunto de métricas que nos permite describir un texto o un grupo de textos. Vamos a aplicar estas medidas al conjunto de canciones para obtener una visión más objetiva.

Seleccionamos 7 métricas básicas:

- **Fernández-Huerta:** índice de legibilidad adaptado al español; valores más altos indican que el texto es más fácil de leer (Wikipedia, 2025c).
- **Gunning Fog:** estima los años de educación formal necesarios para comprender el texto; un valor alto indica mayor complejidad (Wikipedia, 2025e).
- **Automated Readability Index):** calcula la dificultad del texto a partir de la longitud media de las palabras y oraciones; valores altos implican un mayor nivel léxico (Wikipedia, 2025a).
- **SMOG Index:** diseñado para estimar el número de años de educación necesarios para entender un texto (Wikipedia, 2025h).

- **Szigriszt-Pazos:** medida de legibilidad adaptada al español que considera la longitud media de las oraciones y de las palabras; valores altos indican mayor facilidad (Wikipedia, 2025d).
- **Gutiérrez de Polini:** evalúa la dificultad del texto considerando la media de sílabas por palabra; a mayor cantidad de sílabas, mayor complejidad (Wikipedia, 2025f).
- **Crawford:** estima el nivel de comprensión necesario y la duración aproximada de lectura; útil para adaptar textos a diferentes audiencias (Wikipedia, 2025b).

Ahora bien, utilizaremos la letra original de cada canción para aplicar las métricas de textometría seleccionadas. Calcularemos las métricas individualmente para cada canción y, posteriormente, obtendremos un promedio global.

Esto permitirá evaluar si las métricas son coherentes y válidas. Dado que, canciones presentan repeticiones frecuentes, metáforas y símiles que pueden alterar los resultados y generar valores atípicos.

Como sospechábamos, los resultados muestran que varias métricas presentan valores atípicos en su promedio global. Esto confirma que, debido a la naturaleza repetitiva y poética de las canciones, las métricas de legibilidad pueden producir valores poco comunes o extremos.

En la Tabla 3 se presentan los promedios globales obtenidos para cada métrica evaluada.

Cuadro 3: Promedio global de las métricas de textometría.

Métrica	Promedio global
Fernández Huerta	-30.424887
Gunning Fog	62.122454
Automated Readability Index	71.582441
SMOG Index	21.025450
Szigriszt Pazos	-30.854873
Gutiérrez Polini	4.487516
Crawford	3.366969

Estos valores tan altos o negativos en algunos casos reflejan que las canciones, aunque comprensibles para el oyente, no cumplen con los parámetros típicos de textos escritos formales.

Sin embargo, aunque los valores obtenidos no resultan del todo típicos, podríamos analizar las métricas de las canciones a lo largo del tiempo. Para ello, se seguirán varios pasos.

En primer lugar, se realizará una **agrupación temporal** de los promedios por año. Si bien en este conjunto de datos solo hay un valor por año, este procedimiento asegura que el análisis sea escalable en caso de contar, por ejemplo, con el top 10 de canciones por año.

Posteriormente, se aplicará una **normalización MinMax** a cada métrica. Esto evitará que diferencias de escala entre métricas generen distorsiones en la visualización.

Además, para mejorar la interpretación de las tendencias, no se graficará directamente el promedio puntual de cada año. En su lugar, se utilizará una **media móvil de tres años**, lo que suavizará las fluctuaciones extremas y reducirá el impacto de letras particulares que puedan distorsionar el análisis del periodo.

Este enfoque también nos permitirá contrastar nuestra hipótesis: que las canciones antiguas tienden a ser más legibles y poseen una mayor carga semántica en comparación con las composiciones actuales.

La Figura 10 muestra la evolución temporal de las siete métricas seleccionadas, normalizadas y suavizadas mediante una media móvil de tres años.

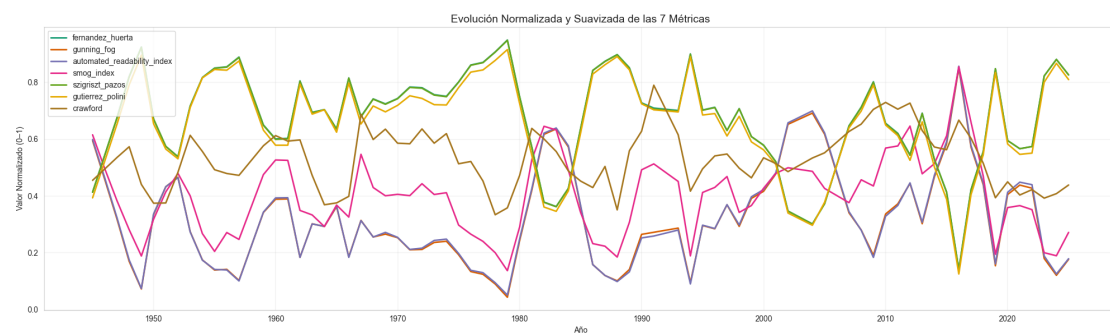


Figura 10: Evolución normalizada y suavizada (media móvil de 3 años) de las 7 métricas de legibilidad a lo largo del tiempo.

En general, las métricas muestran comportamientos irregulares, lo que indica que la interpretación no es sencilla ni concluyente. Además, cualquier inferencia debe realizarse con cautela.

9. Conclusiones

El análisis realizado confirma que las letras de canciones en español, a pesar de su estructura particular con repeticiones y recursos poéticos, siguen en gran medida el comportamiento descrito por la Ley de Zipf. Sin embargo, las métricas de legibilidad aplicadas presentan valores atípicos y comportamientos irregulares, lo que evidencia que deben interpretarse con precaución en este tipo de textos.

La incorporación del análisis fonológico y fonético permitió identificar patrones sonoros recurrentes que podrían influir en la percepción rítmica y melódica de las canciones. En conjunto, los resultados muestran que la combinación de herramientas estadísticas y lingüísticas ofrece una visión más amplia y profunda del lenguaje en la música.

A. Listado de canciones seleccionadas

A continuación, se presenta la lista completa de canciones utilizadas en este estudio, organizadas por año, género, título y artista.

Año	Género	Título	Artista
1945	Porro	<i>Se va el caimán</i>	José María Peñaranda
1946	Bambuco	<i>Soy Colombiano</i>	Rafael Godoy (Garzón y Collazos)
1947	Rumba criolla	<i>La loca Margarita</i>	Hermanas Garavito
1948	Cumbia	<i>Navidad Negra</i>	Los Trovadores de Barú
1949	Foxtrot tropical	<i>Pachito Eché</i>	Alex Tovar (Lucho Bermúdez y su Orq.)
1950	Joropo	<i>Ay si, si</i>	Luis Ariel Rey
1951	Bambuco	<i>Los cisnes</i>	Garzón y Collazos
1952	Vallenato	<i>Alicia la campesina</i>	Los Vallenatos del Magdalena
1953	Porro	<i>El año viejo</i>	Crescencio Salcedo
1954	Porro	<i>La múcura</i>	Lucho Bermúdez y Matilde Díaz
1955	Merecumbé	<i>Cosita linda</i>	Pacho Galán
1956	Vallenato	<i>Lirio rojo</i>	Calixto Ochoa
1957	Rumba criolla	<i>Por vivir en Bogotá</i>	Hermanas Garavito
1959	Cumbia	<i>El pescador</i>	José Barros (Los Trovadores de Barú)
1960	Cumbia	<i>La piragua</i>	José Barros (Los Black Stars)
1961	Cumbia	<i>La pollera colorá</i>	Wilson Choperena (Orq. de P. Salcedo)
1962	Porro	<i>La paloma guarumera</i>	Los Corraleros de Majagual
1963	Vallenato	<i>La casa en el aire</i>	Rafael Escalona (Bovea y sus Vallenatos)

1964	Currulao	<i>A la mina no voy</i>	Leonor González Mina
1965	Bambuco	<i>Pueblito viejo</i>	Garzón y Collazos
1966	Nueva ola	<i>Llorando estoy</i>	Vicky (Esperanza Acevedo)
1967	Cumbia	<i>La piragua</i>	Gabriel Romero & Los Black Stars
1968	Vallenato	<i>Pedazo de acordeón</i>	Alejo Durán
1969	Currulao	<i>Mi Buena Ventura</i>	Peregoyo y su Combo Vacaná
1970	Balada	<i>Llévame contigo</i>	Claudia de Colombia
1971	Cumbia	<i>Los sabanales</i>	Los Corraleros de Majagual (Calixto Ochoa)
1972	Balada pop	<i>Alguien cantó una canción</i>	Billy Pontoni
1973	Salsa	<i>La sirena</i>	Nelson y Sus Estrellas
1974	Salsa	<i>Las caleñas son como las flores</i>	The Latin Brothers (Joe Arroyo)
1975	Salsa	<i>El preso</i>	Fruko y Sus Tesos
1976	Vallenato	<i>La creciente</i>	Binomio de Oro de América
1977	Balada	<i>Llamarada</i>	Isadora (Fernanda Bustos)
1978	Bolero	<i>María de los guardias</i>	Helenita Vargas
1979	Cumbia tropical	<i>Cariñito</i>	Rodolfo Aicardi
1980	Carranga	<i>La cucharita</i>	Jorge Velosa y Los Carrangueros
1981	Joropo	<i>Ay mi llanura</i>	Arnulfo Briceño
1982	Vallenato	<i>Todo es para ti</i>	Diomedes Díaz
1983	Cumbia	<i>Golpe con golpe</i>	Pastor López
1984	Salsa	<i>Cali pachanguero</i>	Grupo Niche
1985	Balada	<i>¿Dónde estará mi primavera?</i>	Raúl Santi
1986	Salsa	<i>Rebelión</i>	Joe Arroyo y La Verdad
1987	Balada	<i>Como un picaflor</i>	Raúl Santi
1988	Rock	<i>La calle</i>	Compañía Ilimitada
1989	Balada	<i>La causa nacional</i>	Sociedad Anónima
1990	Salsa	<i>Una aventura</i>	Grupo Niche

1991	Cumbia	<i>Colombia tierra querida</i>	Lucho Bermúdez & Matilde Díaz
1993	Vallenato fusión	<i>La gota fría</i>	Carlos Vives
1994	Vallenato	<i>El santo cachón</i>	Los Embajadores Vallenatos
1995	Pop latino	<i>Estoy aquí</i>	Shakira
1996	Jazz-folk	<i>Los caminos de la vida</i>	Los Diablitos (Omar Geles)
1997	Tropipop	<i>La tierra del olvido</i>	Carlos Vives
1998	Rock	<i>Florecita rockera</i>	Aterciopelados
1999	Pop	<i>Me voy</i>	Andrés Cepeda
2000	Rock	<i>Fíjate bien</i>	Juanes
2001	Pop	<i>Suerte (Whenever, Wherever)</i>	Shakira
2002	Rock	<i>A Dios le pido</i>	Juanes
2004	Salsa	<i>La pantera mambo</i>	La 33
2005	Pop	<i>La camisa negra</i>	Juanes
2007	Hip-hop	<i>Somos pacífico</i>	ChocQuibTown
2008	Electrocumbia	<i>Fuego</i>	Bomba Estéreo
2009	Electro-folk	<i>Mi Kolombia</i>	Systema Solar
2010	Pop	<i>Waka Waka</i>	Shakira (ft. Freshly-ground)
2011	Hip-hop	<i>Chocolate</i>	Profetas
2012	Swing	<i>Suín Romanticón</i>	Monsieur Periné
2013	vallenato	<i>Volví a nacer</i>	Carlos Vives
2014	Reguetón	<i>6 AM</i>	J Balvin (ft. Farruko)
2015	Reguetón	<i>Ginza</i>	J Balvin
2016	Vallenato	<i>La bicicleta</i>	Carlos Vives & Shakira
2017	Reguetón	<i>Mi gente</i>	J Balvin & Willy William
2018	Reguetón	<i>Mi cama</i>	Karol G
2019	Reguetón	<i>Tusa</i>	Karol G (ft. Nicki Minaj)
2020	Reguetón	<i>Hawái</i>	Maluma
2021	Reguetón	<i>Bichota</i>	Karol G
2022	Reguetón	<i>Provenza</i>	Karol G
2023	Reguetón	<i>TQG (Te Quedó Grande)</i>	Karol G & Shakira

2024	Reguetón	<i>LUNA</i>	ATL Jacob & Feid
2025	Pop	<i>Enamorarte Mil Veces</i>	Fonseca & Manuel Medrano

Referencias

- El Tiempo. (2012). *Las 50 mejores canciones de Colombia* [Consultado el 9 de agosto de 2025]. <https://www.eltiempo.com/don-juan/cultura/las-50-mejores-canciones-de-colombia+articulo+12683827>
- FULANITOVIAJERO Periodista Musical. (2023). *Historia musical de Colombia* [Consultado el 9 de agosto de 2025]. https://www.youtube.com/watch?v=mUyhGhySNek&list=RDmUyhGhySNek&start_radio=1
- Radio Nacional de Colombia. (2014). *Radio Nacional de Colombia: 80 años en 80 canciones* [Consultado el 9 de agosto de 2025]. <https://www.radionacional.co/cultura/radio-nacional-de-colombia-80-anos-en-80-canciones>
- Wikipedia. (2025a). *Automated readability index* [Consultado el 10 de agosto de 2025]. https://en.wikipedia.org/wiki/Automated_readability_index
- Wikipedia. (2025b). *Crawford's readability formula* [Consultado el 10 de agosto de 2025]. https://en.wikipedia.org/wiki/Readability#Crawford's_formula
- Wikipedia. (2025c). *Fernández-Huerta readability formula* [Consultado el 10 de agosto de 2025]. https://es.wikipedia.org/wiki/F%C3%B3rmula_de_lecturabilidad_de_Fern%C3%A1ndez_Huerta
- Wikipedia. (2025d). *Fórmula de Szigriszt Pazos* [Consultado el 10 de agosto de 2025]. https://es.wikipedia.org/wiki/F%C3%B3rmula_de_Szigriszt_Pazos
- Wikipedia. (2025e). *Gunning fog index* [Consultado el 10 de agosto de 2025]. https://en.wikipedia.org/wiki/Gunning_fog_index
- Wikipedia. (2025f). *Índice de legibilidad de Gutiérrez de Polini* [Consultado el 10 de agosto de 2025]. https://es.wikipedia.org/wiki/%C3%8Dndice_de_legibilidad_de_Guti%C3%A9rrez_de_Polini
- Wikipedia. (2025g). *Ley de Zipf* [Consultado el 9 de agosto de 2025].
- Wikipedia. (2025h). *SMOG grade* [Consultado el 10 de agosto de 2025]. <https://en.wikipedia.org/wiki/SMOG>