

<https://doi.org/10.32735/S0718-22012026000624195>

79-94

PATRONES CUANTITATIVOS EN POEMAS CHILENOS DE LOS SIGLOS XIX A XXI

Quantitative patterns in Chilean poems between the 19th and 21st centuries

RICARDO MARTÍNEZ-GAMBOA

Universidad Diego Portales (Chile)

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

ricardo.martinez@udp.cl

VALERIA ARAYA-LOPETEGUI

Universidad Diego Portales (Chile)

<https://orcid.org/0000-0002-2845-6427>

valeriaant.araya@mail.udp.cl

DANIEL MORA MELANCHTHON

Pontificia Universidad Católica de Valparaíso (Chile)

<https://orcid.org/0000-0002-2758-5198>

daniel.mm91@gmail.com

NICOLÁS ACOSTA

Universidad Nacional de Cuyo (Argentina)

<https://orcid.org/0000-0003-1140-2964>

nmac.1996@gmail.com

SUYLI APROSIO BARICIC

Universidad Diego Portales (Chile)

<https://orcid.org/0009-0004-8138-0750>

suyli.aprosio@gmail.com

Resumen

Se presenta un análisis cuantitativo estilométrico de poesía chilena a partir de un corpus especializado, el Corpus de Poesía Chilena (Corpoch), un enfoque que ha experimentado un notable desarrollo desde los trabajos de Franco Moretti (Goodwin y Holbo, 2011; Moretti, 2005). Este proyecto, que, de forma inédita, dispone de un corpus sistemático, surge con el propósito de proporcionar a las y los investigadores en poesía chilena tendencias, patrones, regularidades y excepciones sustentadas en evidencia robusta a través de rutinas computacionales (Yu, 2008; Marchetti, Sprugnoli y Tonelli, 2014) que aplican los avances de la Lingüística Computacional y de Corpus al estudio de la poesía chilena producida entre los siglos XIX y XXI.

Palabras clave: Corpus; lingüística computacional; poesía chilena; patrones métricos; demografía de poetas de Chile.

Recibido: 17 enero 2025

Aceptado: 4 junio 2025

Abstract

A quantitative stylometric analysis of Chilean poetry is presented based on a specialized corpus, *Corpus de Poesía Chilena* (Corpoch), an approach that has advanced significantly since the foundational work of Franco Moretti (Goodwin and Holbo, 2011; Moretti, 2005). This project –unprecedented in providing a systematic corpus– aims to furnish scholars of Chilean poetry with trends, patterns, regularities, and exceptions grounded in robust evidence through computational routines (Yu, 2008; Marchetti, Sprugnoli and Tonelli, 2014) that apply recent advances in Computational and Corpus Linguistics to poetry produced in Chile between the nineteenth and twenty-first centuries.

Keywords: Corpus; computational linguistics; Chilean poetry; metric patterns; demographics of Chilean poets.

1. INTRODUCCIÓN

El análisis cuantitativo de la literatura, definido como un método discursivo y estadístico, es una herramienta clave en la investigación literaria contemporánea, lo que ha ofrecido perspectivas diferentes que trascienden los enfoques tradicionales. Inicialmente, con el desarrollo del concepto de “lectura distante” introducido por Franco Moretti (Goodwin y Holbo, 2011; Moretti, 2005), el procesamiento masivo de datos textuales ha permitido desentrañar regularidades y tendencias en los textos literarios que, de otro modo, permanecerían invisibles a través de la exégesis individual. Este enfoque metodológico, vinculado a los avances en Lingüística Computacional y Lingüística de Corpus, replantea el análisis literario al combinar herramientas tecnológicas con principios de las humanidades digitales, proporcionando una base sólida para estudios literarios que integran precisión estadística y sensibilidad estética.

En el ámbito anglosajón, en particular en la última década, ha habido abundante investigación sobre propiedades cuantitativas de poemas mediante rutinas computacionales (Yu, 2008; Šel, Plecháč y Lassche, 2022; Marchetti et al., 2014). Revistas especializadas como *Digital Scholarship in the Humanities* exploran herramientas y análisis lingüísticos ligados a las Humanidades Digitales. Si bien para la poesía escrita en lengua española existen estudios y modelos de análisis automatizado, en particular, la lírica chilena ha sido poco estudiada desde la perspectiva de corpus o Lingüística Computacional (Hernández, De Sisto, Pérez, De la Rosa, Ros y González-Blanco, 2021; De la Ros, Pérez, Hernández, Ros y González-Blanco, 2020; Anttila y Heuser, 2016; Gervás, 2000, 2001).

A raíz de que la crítica chilena ha privilegiado usualmente los abordajes cualitativos del estudio de la poesía, capaces de captar matices semánticos y contextuales, el proyecto que se presenta emerge de la necesidad de desarrollar una mirada cuantitativa que permita escrutar con criterios reproducibles de textos pertenecientes a un corpus de poesía chilena. Todo esto con el fin de detectar tendencias, patrones y excepciones en las

características formales y semánticas de los versos. El objetivo es ofrecer a la comunidad académica información basada en evidencia robusta para el análisis de la poesía chilena, enfocándose en las características métricas y rítmicas de los poemas y sus propiedades emocionales, reconocidas en estudios hermenéuticos de la lírica.

El Corpus de la Poesía Chilena (Corpoch) cubre poesía chilena de entre los siglos XIX y XXI, esto es, desde los inicios de la República hasta hoy. Incluye información de 2.244 obras registradas en el catastro de la Biblioteca Nacional de Chile, firmadas por 1.107 autores y autoras. Adicionalmente a esto, el corpus también recoge información demográfica de otras bases de datos abiertas. Además, incluye información correspondiente a la edad, generación, filiaciones, género y procedencia de las y los poetas incluidos. De esta forma, este corpus permite revisar con criterios estadísticos regularidades que pueden ser invisibles a la lectura manual. Se propone, a través del uso de esta información, un análisis sistemático y cuantitativo de un corpus representativo de poesía chilena.

El diseño del corpus ha contemplado la inclusión de obras ampliamente reconocidas y, también, aquellas que reflejan las diversidades topográficas, generacionales y estilísticas de la poesía chilena. Lo anterior permite incorporar variables demográficas como género, etnia y filiación estética. Para ello, se han implementado herramientas computacionales avanzadas capaces de identificar patrones formales y semánticos en los versos, como la segmentación métrica, la distribución de patrones rítmicos y la categorización semántica basada en algoritmos lingüísticos robustos (Hernández et al., 2021).

Mediante este enfoque, el presente estudio ha identificado tendencias y excepciones en las características formales de la poesía chilena para proporcionar un recurso metodológico que sirva de base para futuras investigaciones en el campo de los estudios literarios hispanoamericanos. Se trata de análisis estructurado a partir de una metodología cuantitativa y computacional que se articula en torno a un doble objetivo: por un lado, ampliar el conocimiento existente sobre la tradición poética chilena desde una perspectiva renovada, y, por otro, contribuir al desarrollo de las humanidades digitales en el ámbito de la literatura en español. De este modo, el trabajo no solo aborda preguntas fundamentales sobre la evolución de las formas poéticas en Chile, sino que también establece un precedente en la aplicación de métodos estadísticos y computacionales al estudio de la lírica, abriendo nuevas posibilidades para el análisis literario en el siglo XXI.

2. DE CORPUS DE REFERENCIA A CORPUS ESPECIALIZADOS

El desarrollo del procesamiento del lenguaje natural (PLN) ha transformado los estudios lingüísticos y literarios, permitiendo abordar preguntas sobre estructuras y usos del lenguaje desde perspectivas empíricas y cuantitativas. El primer corpus digital

desarrollado fue el *Corpus of Present-Day American English* de la Universidad de Brown (Francis y Kučera, 1961). En castellano, actualmente, se encuentran diversos corpus de referencia, entre los que destacan el *Corpus del Español* o los construidos por la Real Academia Española (Davies, 2002; RAE, 1994, 2001, 2014).

En el caso de los corpus especializados en poesía, también encontramos ejemplos destacados como el *Gutenberg Project* (Hart, 1971), el *Corpus de la antigua lírica popular hispánica* (Alatorre, 2003), el *Corpus de la poesía castellana del Siglo de Oro* (Navarro-Colorado et al., 2015) y el *PoetryLab* (POSTDATA, 2020). Estos corpus presentan características distintivas –como segmentación silábica, conteo de sílabas, licencias métricas y análisis de estructuras rítmicas– que permiten identificar patrones estilísticos y métricos con un nivel de granularidad que difícilmente podría alcanzarse con métodos tradicionales. Este tipo de enfoque facilita la identificación de regularidades que enriquecen la comprensión tanto del lenguaje especializado como de la estructura poética.

Para el ámbito chileno, en consecuencia, se ha generado para este estudio el *Corpoch*, que se configura como un corpus especializado que integra tanto datos textuales como variables demográficas, incluyendo género, generación literaria, región de origen y filiación estética. Con más de 15.000 poemas y 400.000 versos, este corpus permite realizar análisis cuantitativos de características formales y estilísticas, como patrones métricos y acentuales, así como correlaciones entre variables contextuales y lingüísticas. Lo anterior aporta datos sólidos para el estudio de la poesía chilena y establece un precedente en la sistematización de corpus poéticos en lengua española, superando las limitaciones de los estudios individuales y ofreciendo una herramienta para futuras investigaciones.

3. EL ANÁLISIS CUANTITATIVO DE LA LITERATURA

El análisis cuantitativo de la literatura se ha consolidado como una metodología que combina enfoques estadísticos y computacionales para explorar aspectos formales, estilísticos y semánticos en los textos literarios, asociado a la “lectura distante” propuesta por Franco Moretti (2005) que ha revolucionado el campo de los estudios literarios al incorporar herramientas propias de las humanidades digitales y la Lingüística Computacional. La lectura distante, como enfoque, se aleja del análisis manual de textos individuales para centrarse en patrones globales que emergen al estudiar amplios conjuntos de obras, lo que permite identificar regularidades que enriquecen la comprensión de los fenómenos literarios desde una perspectiva cuantitativa y empírica.

Esta metodología se relaciona, en el caso del presente estudio, con el PLN, una disciplina que comenzó a desarrollarse en la década de 1960 con el propósito de sistematizar datos lingüísticos (Moreno Sandoval, 1998). Desde entonces, la Lingüística de Corpus y la Lingüística Computacional han proporcionado herramientas avanzadas para procesar grandes volúmenes de datos textuales, lo que facilita el análisis de aspectos como la frecuencia léxica, las estructuras sintácticas y las propiedades semánticas. En el

contexto literario, estas técnicas han permitido abordar preguntas complejas sobre la evolución de los estilos literarios, la influencia de contextos históricos en las obras, y las dinámicas de producción y recepción literaria.

Uno de los aportes más destacados de la lectura distante es su capacidad para abordar fenómenos de macroevolución literaria. Como señala Moretti (2005), el análisis cuantitativo permite observar patrones que no son evidentes a través de la lectura tradicional, como las correlaciones entre géneros literarios y periodos históricos o las preferencias estilísticas que caracterizan generaciones de escritores (Goodwin y Holbo 2011). En este sentido, el análisis cuantitativo se articula como una metodología y, a su vez, como un marco epistemológico que plantea nuevas preguntas para los estudios literarios.

Entre las metodologías destacadas en el análisis cuantitativo, el modelo biberiano se ha consolidado como una herramienta clave en el estudio de textos. Propuesto por Biber (1988), este modelo evalúa las características gramaticales y semánticas de los textos para clasificarlos en categorías según sus dimensiones funcionales. Estas dimensiones incluyen, por ejemplo, el grado de formalidad o informalidad, así como la orientación narrativa o descriptiva de un texto. El modelo no solo permite identificar patrones lingüísticos específicos, sino que también agrupa textos con funciones similares en un mismo espacio interpretativo, como cartas al director y artículos de opinión, o mensajes informales como chats y correos laborales. Este enfoque, basado en el análisis estadístico de las características lingüísticas, ha sido fundamental para abordar preguntas sobre el estilo y la funcionalidad textual.

En el ámbito hispanoamericano, los estudios cuantitativos han sido menos frecuentes que en contextos anglosajones. Sin embargo, investigaciones como las realizadas sobre narrativa chilena utilizando el modelo biberiano ya mencionado han demostrado el potencial de estas técnicas para identificar dimensiones lingüísticas y estilísticas en conjuntos de obras. Por ejemplo, trabajos previos (Martínez-Gamboa, 2015, 2016) han aplicado análisis factorial y algoritmos de agrupamiento o *clustering* para reunir textos según propiedades lingüísticas compartidas, evidenciando cómo la literatura refleja patrones inherentes tanto a nivel temático como formal.

En el caso de la poesía, el análisis cuantitativo permite explorar fenómenos únicos de este género, como las estructuras métricas, las regularidades rítmicas y la semántica de los versos. Estudios internacionales han utilizado herramientas como el *Linguistic Inquiry and Word Count* (LIWC) (Pennebaker, Francis y Booth, 2001) para analizar la distribución de categorías semánticas en poemas y establecer correlaciones entre aspectos estilísticos y emocionales. Sin embargo, en español, este tipo de análisis ha sido más limitado, subrayando la necesidad de proyectos que aborden estas dinámicas de manera sistemática y representativa.

El análisis cuantitativo de la poesía chilena, como se plantea en este proyecto, busca superar estas carencias al aplicar metodologías avanzadas para identificar tendencias, patrones y excepciones en un corpus demográficamente representativo. Este enfoque complementa los estudios hermenéuticos tradicionales y amplía de manera significativa el horizonte de los estudios literarios al introducir datos empíricos que fundamentan las interpretaciones y abren nuevas posibilidades para comprender la complejidad de la producción poética.

4. METODOLOGÍA

Metodológicamente, el presente estudio tiene un enfoque cuantitativo y computacional para el análisis de un corpus representativo de poesía chilena que abarca obras producidas entre los siglos XIX y XXI. La metodología empleada integra principios de la Lingüística de Corpus y la Lingüística Computacional combinando herramientas de PLN con técnicas de análisis estadístico avanzadas. Este enfoque permite identificar patrones léxicogramaticales, métricos y semánticos en los versos, así como establecer correlaciones entre las características formales de los textos y variables contextuales como género, región y generación literaria.

4.1. DISEÑO DEL CORPUS

El Corpoch incluye una muestra diversa de 15.156 poemas escritos por 1.107 poetas de nacionalidad chilena. La selección de textos se basó en criterios de representatividad demográfica, considerando variables como el lugar de nacimiento, el sexo, la etnia y la generación literaria. Adicionalmente, se incluyeron datos contextuales, como premios literarios, estudios realizados y lugar de fallecimiento, lo que permite un análisis integrado de los factores lingüísticos y sociohistóricos.

El procesamiento inicial del corpus implicó una etapa de limpieza y normalización de los datos, eliminando inconsistencias en la transcripción de textos y asegurando la compatibilidad de los archivos con las herramientas computacionales empleadas. El formato estándar adoptado para los textos fue el XML/TEI (Barcala, Blanco y Darriba, 2006), que facilita la anotación de metadatos y la segmentación de unidades textuales como versos (como la señalada por Aguirrezabal, 2017), estrofas (De la Rosa, Pérez, Hernández Lorenzo, De Sisto, Ros y González-Blanco, 2021) y poemas completos.

4.2. TÉCNICAS DE ANÁLISIS

Para el desarrollo del análisis se utilizaron tres elementos. En primer lugar, se trabajó desde el PLN, en donde se emplearon herramientas como LIWC y modelos de análisis léxico basados en Python para identificar frecuencias léxicas y categorías semánticas predominantes en el corpus. Estas herramientas permiten categorizar

términos según dominios emocionales, conceptuales y estilísticos, y ofrecen una visión granular de los elementos semánticos en los textos.

En segundo lugar, se trabajó con el análisis métrico y rítmico. Se implementaron algoritmos que detectan patrones rítmicos y licencias métricas, basados en reglas de versificación castellana. Estos algoritmos fueron adaptados de modelos desarrollados por Navarro-Colorado (2017), y este junto a las colaboradoras Ribes Lafoz y Sánchez (2015, 2016), y optimizados para el análisis de poesía chilena. Se realizaron análisis de conteo silábico, distribución de acentos y clasificación de rimas.

Por último, se realizó un análisis estadístico en el que los datos extraídos del corpus fueron procesados mediante técnicas de análisis factorial y *clustering* para identificar agrupaciones textuales basadas en similitudes métricas, léxicas y semánticas. Adicionalmente, se utilizaron modelos de regresión para explorar correlaciones entre las características formales de los poemas y variables contextuales, como la época de producción y la filiación estética de las y los autores.

4.3. FASES DEL ESTUDIO Y LIMITACIONES

El estudio se inició con la construcción del corpus. Así, se realizó una recolección, selección y digitalización de los diferentes textos poéticos y sus metadatos. Luego, se realizó el procesamiento y análisis computacional. En esta etapa, se aplicaron técnicas de PLN y el análisis de la métrica que permitió extraer datos ligados a la estructura y formación de elementos del corpus. Finalmente, se interpretaron y validaron los datos a la luz de la teoría, y se compararon los hallazgos con estudios anteriores por medio de la identificación de tendencias realizadas con pruebas estadísticas.

Un elemento importante que debe considerarse es que el corpus incluye textos protegidos por derechos de autor, por lo que su acceso está restringido a fines de investigación académica. Los resultados obtenidos están disponibles en forma de datos agregados y estadísticas descriptivas para garantizar el cumplimiento de las normativas legales vigentes. Además, se reconocen las limitaciones inherentes al análisis computacional, en particular en lo relativo a la interpretación semántica, la que puede verse influenciada por las particularidades culturales y estilísticas de los textos poéticos.

5. RESULTADOS

El Corpoch es una muestra representativa de la producción poética chilena que integra 15.156 poemas escritos por 1.107 poetas, lo que equivale a más de 400.000 versos y aproximadamente 1.2 millones de palabras. Este diseño fue concebido para garantizar la diversidad y amplitud necesarias para un análisis robusto de tendencias y excepciones en las características formales y semánticas de los versos.

Tabla 1. Estadísticas generales (cantidad de poetas, poemas, versos y palabras representadas):

Dato	Cantidad
Poetas	1.107
Poemas	15.058
Estrofas	78.446
Versos	408.210
Palabras	2.538.252

El corpus incluye información complementaria sobre los y las autoras, como género, generación literaria, región de origen y premios literarios. Estos datos proporcionan un marco contextual que facilita el análisis de patrones estilísticos en relación con las dinámicas históricas y sociales. Por ejemplo, el corpus refleja un aumento progresivo en la participación femenina, especialmente desde la segunda mitad del siglo XX, lo que subraya cambios significativos en la inclusión de voces poéticas diversas.

Tabla 2. Variables contextuales (género, generación literaria, premios).

Variable	Categorías	Cantidad de Poetas	Porcentaje (%)
Género	Masculino	720	65%
	Femenino	387	35%
Generación Literaria	Siglo XIX	250	22.6%
	Siglo XX (1900-1950)	420	38%
	Siglo XX (1951-2000)	300	27.1%
	Siglo XXI	137	12.3%
Premios	Sin premios registrados	900	81.3%
	Premio Nacional de Literatura	45	4.1%
	Premios Internacionales	50	4.5%
	Otros reconocimientos	112	10.1%

Continuando con los análisis extraídos a partir del corpus, se comparó la cantidad de poetas de una misma generación literaria en dos regiones del país: la del Libertador General Bernardo O'Higgins y la del Biobío. Los resultados evidencian una predominancia de poetas de la generación de 1942 en la primera y de 1987 en la segunda, lo que refleja una diferencia en la publicación de material poético en distintas zonas del territorio nacional, revelando así diversas realidades socioculturales según la época y la región.

Tabla 3. Generación literaria con mayor cantidad de poetas en dos regiones

Región de origen	Generación con más poetas
Región del Libertador General Bernardo O'Higgins	1942
Región del Biobio	1987

Asimismo, se examinó el porcentaje de poetas mujeres nacidas antes y después de la década de 1950. El corpus muestra que un 37,9 % de las poetas chilenas nació antes o

durante el año 1950, mientras que el 62,1 % nació después, demostrando un creciente interés y desempeño nacional en el área de la poesía a lo largo del siglo XX. Los datos obtenidos del cruce de variables como región de origen, fecha de nacimiento o generación literaria permiten un análisis cuantitativo con múltiples combinaciones posibles. Esto evidencia el amplio potencial de este tipo de corpus para investigaciones futuras.

Tabla 4. Porcentaje de poetas del siglo XX nacidos/as antes y después de 1950

Fecha de nacimiento	%
Hasta 1950	37,9
Desde 1950	62,1

El análisis formal del corpus revela una clara predominancia de estructuras métricas tradicionales, como el endecasílabo (15,14 %) y el octosílabo (13,07 %), junto con un aumento significativo del verso libre en las producciones del siglo XX. Estas tendencias dan cuenta de una transición desde formas poéticas clásicas hacia estructuras más experimentales en el marco de las Vanguardias.

Tabla 5. Porcentaje de versos según cantidad de sílabas

Sílabas	Número	%
2	6840	1,68
3	10347	2,54
4	14642	3,60
5	17195	4,23
6	16950	4,17
7	38533	9,48
8	53133	13,07
9	35359	8,70
10	23882	5,87
11	61565	15,14
12	24483	6,02
13	20106	4,95
14	26030	6,40
+14	57523	14,15

De la misma manera, es posible estudiar las estructuras métricas predominantes en los versos de cada poeta en particular. A modo de ejemplo, el corpus evidenció la predilección de Violeta Parra por los versos octosílabos, y de Nicanor Parra por los endecasílabos. La tendencia hallada sugiere una conexión con la métrica tradicional, utilizada en mayor cantidad por poetas chilenos de la época, tal como se demostró en el resultado anterior. A partir de estos datos, en el caso de ambos poetas, se podría estudiar una posible intencionalidad política al utilizar las estructuras métricas más populares del país durante la época del siglo XX.

Tabla 6. Tipo de verso más utilizado en los poemas de Violeta Parra y Nicanor Parra

	Poemas	Versos	Verso preferido
Violeta Parra	218	9585	Octosílabo
Nicanor Parra	740	21940	endecasílabo

Sin embargo, pese a los cambios hacia estructuras más vanguardistas, el análisis de las palabras más utilizadas al final de los versos permite constatar una continuidad en las temáticas universales y tradicionales de la poesía: la vida, la muerte, el amor y la naturaleza. De acuerdo con ambos estudios anteriores, se advierte una intención de renovar la métrica y la estructura visual del poema, sin abandonar los motivos líricos fundamentales. Esta coexistencia de innovación formal y persistencia temática se ajusta a la percepción de la época como un periodo de transición hacia la Vanguardia.

Tabla 7. Frecuencia de palabras utilizadas al final de los versos escritos por poetas chilenos y chilenas del siglo XX

Palabras finales de versos	#
1 tierra	2035
2 vida	1775
3 noche	1712
4 muerte	1466
5 ojos	1463
6 mundo	1393
7 día	1349
8 cielo	1321
9 tiempo	1271
10 mar	1180
11 amor	1094
12 silencio	1081
13 nada	1079
14 viento	1033
15 agua	993

En términos de rima, los esquemas consonantes prevalecen en los textos anteriores a 1900, mientras que los versos blancos y las rimas asonantes dominan en las producciones posteriores, en particular en las obras modernas y contemporáneas.

Tablas 8 y 9. Frecuencia de tipos de verso y estructuras métricas predominantes

Tipo de estrofa	Casos	%	Tipo específico Métrica versos	Número	%
Estrofa libre	7132	16,46	Octosílabo trocaico	9689	9,10
Couplet	6004	13,86	Heptasílabo Dactílico	8909	8,36
Terceto	3397	7,84	Trisílabo Anfibráquico	8815	8,27
Serventesio	2813	6,49	Endecasílabos A maiore (A) melódico	8296	7,79

Sexteto	2404	5,55	Endecasílabos A maiore (A) heróico	5768	5,41
Quinteto	2002	4,62	Bisílabo Trocaico	5719	5,37
Septeto	1622	3,74	Hexasílabo Dactílico	5193	4,87
Romance arte mayor	1467	3,39	Octosílabo Mixto A	4702	4,41
Quintilla	1396	3,22	Pentasílabo Trocaico	4475	4,20
Novena	1336	3,08	Pentasílabo Dactílico o Adónico	4180	3,92
Octava	1220	2,82			
Silva arromanzada	1166	2,69			
Seguidilla	1121	2,59			
Cantar	980	2,26			
Espinela	811	1,87			

Sin embargo, los análisis individuales de ciertos autores permitieron identificar excepciones significativas que enriquecen el panorama general. Un caso emblemático es el de Pablo Neruda, donde se observó una evolución en la extensión de los versos de sus poemas. A lo largo de su obra, los versos tienden a acortarse en la medida en que sus producciones avanzan en el tiempo. Este hallazgo sugiere un posible cambio estilístico y temático en su escritura. Dicho fenómeno podría ser representativo de una búsqueda de síntesis expresiva y una adaptación a los movimientos literarios contemporáneos a su obra tardía. En futuros estudios, este análisis podría replicarse en el trabajo de más poetas relevantes, como Gabriela Mistral, Vicente Huidobro, Raúl Zurita y otros, de manera que, una vez que sus obras puedan ser organizadas y codificadas cronológicamente en la base de datos, sea posible realizar comparaciones en la evolución métrica de sus escrituras.

Tabla 10. Frecuencias de la extensión promedio de versos en la obra de Pablo Neruda

Obra	Año de publicación	Extensión promedio de los versos
Crepusculario	1923	14
Residencia en la Tierra	1933	12
Canto General	1950	10
Estravagario	1958	8

A nivel semántico, se pueden encontrar diversos resultados basados en herramientas de Lingüística Computacional. En ese sentido, se identificó que las palabras más frecuentes en el corpus se agrupan en categorías temáticas como naturaleza (*mar, tierra, cielo*), emociones (*amor, muerte, dolor*) y temporalidad (*vida, tiempo, noche*). Estas categorías reflejan los ejes temáticos predominantes de la poesía chilena a lo largo de su historia.

Un análisis diacrónico muestra que los términos relacionados con emociones negativas (*muerte, dolor*) aumentaron durante el siglo XX, lo que refleja un contexto de crisis sociopolítica. Por otro lado, las palabras asociadas a la naturaleza fueron más frecuentes en el siglo XIX, en línea con las influencias románticas de la época.

Un elemento importante tiene que ver con el análisis de la carga emocional de los textos poéticos; en ese sentido, resulta relevante mencionar, a modo preliminar, que se

está revisando la posible correlación entre la medida del verso y la polaridad positiva o negativa a nivel semántico. Este punto resulta relevante tanto para conocer el fenómeno de la métrica en relación con los temas trabajados como para revisar las posibles excepciones, puesto que esto puede dar luz sobre giros particulares, contradicciones o decisiones estilísticas.

Tabla 11. Palabras frecuentes por categorías temáticas

Categoría	Palabras más frecuentes	Frecuencia absoluta	Frecuencia relativa (%)
Naturaleza	Mar, tierra, cielo, sol, luna, agua, montaña	8,950	15%
Emociones	Amor, muerte, dolor, alegría, tristeza, vida	7,500	12%
Temporalidad	Tiempo, noche, día, eternidad, instante, ayer	6,800	11%

Un elemento importante tiene que ver con el uso de técnicas de *clustering* y análisis factorial. Esto permitió identificar grupos estilísticos en el corpus, donde los poemas se agruparon según similitudes métricas y semánticas. Por ejemplo, se observó que los poemas con un alto grado de experimentación métrica tienden a correlacionarse con autoras y autores de generaciones vanguardistas y posmodernas. Este punto es fundamental para poder hacer un recorrido sobre las tendencias de diferentes épocas y establecer características generacionales o de grupos específicos a nivel diatópico o diacrónico.

Además, el análisis destacó excepciones notables, como obras que combinan estructuras tradicionales con elementos léxicos que resultan innovadores en su periodo, lo que subraya la complejidad estilística de algunas y algunos poetas y los múltiples usos que puede tener el trabajo con un corpus tan grande para el desarrollo de análisis de las formas y de los temas.

Tabla 12. Agrupación de poemas según análisis factorial y *clustering*

Grupo	Características principales	Ejemplo de poemas	Cantidad de poemas	Proporción
Grupo 1	Estructuras métricas tradicionales, rima consonante, temática clásica (naturaleza, amor).	<i>Crepusculario</i> (P. Neruda)	4,200	28%
Grupo 2	Verso libre, léxico abstracto, influencia vanguardista.	<i>Residencia en la Tierra</i> (P. Neruda)	3,600	24%
Grupo 3	Temática social, léxico emocional negativo, extensiones variadas.	<i>Canto General</i> (P. Neruda)	2,800	19%
Grupo 4	Lenguaje simbólico, temática mística o filosófica.	<i>Altazor</i> (V. Huidobro)	2,400	16%
Grupo 5	Estructuras experimentales, temas contemporáneos (identidad, globalización)	<i>Estravagario</i> (P. Neruda)	2,156	13%

6. CONCLUSIONES

El presente estudio ha permitido explorar, desde una perspectiva cuantitativa y computacional, las características formales y semánticas de la poesía chilena producida entre los siglos XIX y XXI. Mediante la aplicación de herramientas de Lingüística Computacional y Lingüística de Corpus, los resultados obtenidos ofrecen un panorama amplio y detallado que contribuye al análisis e interpretación de esta rica tradición poética.

En primer lugar, la descripción general del corpus destaca su amplitud y representatividad, abarcando más de 15.000 poemas de 1.107 poetas, con variables demográficas que reflejan las dinámicas históricas y sociales de la producción poética chilena. Este diseño permite desarrollar análisis formales, estructurales y semánticos rigurosos, además de facilitar la contextualización histórica de los hallazgos. Al integrar variables como género, generación literaria y región de origen, el Corpoch permite documentar las tendencias literarias en distintos periodos del Chile republicano y dilucidar preferencias estilísticas y temáticas de las y los autores a lo largo de los años.

Desde el punto de vista formal y estructural, los resultados métricos revelan una transición significativa desde las estructuras tradicionales, dominadas por endecasílabos y rimas consonantes, hacia formas experimentales y verso libre, sobre todo en el siglo XX. Estos cambios reflejan la influencia de movimientos estéticos como las Vanguardias y la posmodernidad, y subrayan una diversificación estilística que enriquece el panorama poético chileno. Sin embargo, se mantiene el uso métrico en la poesía en algunos y algunas autoras, lo que complejiza y enriquece el proceso de análisis añadiendo las excepciones de cada tradición específica.

En el ámbito semántico, se identificaron categorías temáticas recurrentes, como la naturaleza, las emociones y la temporalidad con variaciones en la prevalencia de estos temas a lo largo de los siglos. En el siglo XIX se muestra una inclinación hacia la exaltación de la naturaleza, mientras que en el XX predominan temas asociados a la crisis y el desarraigo, en sintonía con el contexto sociopolítico del país. Este punto es en particular interesante, porque esto se puede cruzar con el elemento histórico para visualizar las relaciones que se establecen entre las estructuras y temas, y las vivencias del contexto de producción.

Tomando los hallazgos demográficos y métricos documentados, el Corpoch se vuelve una herramienta valiosa para el estudio de la poesía chilena. Las investigaciones, antes limitadas por la falta de un corpus sistematizado, ahora podrán explorar nuevos aspectos y servir como base para estudios más exhaustivos. Al igual que proyectos destacados como el *Nuevo corpus de la antigua lírica popular hispánica* (Alatorre, 2003) y el *Diachronic Spanish Sonnet Corpus (DISCO)* (Navarro-Colorado et al., 2016), el Corpoch proporciona un volumen significativo de datos que facilitan tanto análisis lingüísticos como interpretaciones culturales.

Este corpus valida la importancia de integrar herramientas computacionales en el análisis literario y constituye un punto de partida para futuras investigaciones. Los hallazgos refuerzan la necesidad de usar tecnologías para identificar patrones formales y semánticos, enriqueciendo a los estudios literarios como los enfoques tradicionales en lingüística.

Esta investigación se ha realizado en el marco del Proyecto ANID Fondecyt Iniciación N° 11230032.

OBRAS CITADAS

- Aguirrezabal, Manex (2017). *Automatic Scansion of Poetry* (Tesis de doctorado). University of the Basque Country.
- Alatorre, Margit Frenk (2003). *Nuevo corpus de la antigua lírica popular hispánica, siglos XV a XVII*. (Vol. 1). UNAM.
- Anttila, Arto y Heuser, Ryan (2016). Phonological and Metrical Variation across Genres. *Proceedings of the Annual Meetings on Phonology*, 3.
- Barcala, Francisco; Blanco, Cristina; y Darriba, Víctor (2006). Metodología para la construcción de corpóra textuales estructurados basados en XML. *Procesamiento del Lenguaje Natural*, 36.
- Biber, Douglas (1988). *Variation across Speech and Writing*. Cambridge University Press.
- Davies, Mark (2002). *Corpus del español: 100 million words, 1200s-1900s*. <https://www.corpusdelespanol.org/x.asp?c=2>
- De la Rosa, Javier; Pérez, Álvaro; Hernández, Laura; Ros, Salvador; y González-Blanco, Elena (2020). Rantanplan, Fast and Accurate Syllabification and Scansion of Spanish Poetry. *Procesamiento del Lenguaje Natural*, 65, 83-90.
- De la Rosa, Javier; Pérez, Álvaro; Hernández, Laura; De Sisto, Mirella; Ros, Salvador; y González-Blanco, Elena (2021). *A Hybrid Approach to Stanza Classification in Spanish Poetry* [sesión de conferencia]. EADH 2021, Krasnoyarsk, Rusia.
- Francis, W. Nelson y Kučera, Henry (1961). *A Standard Corpus of Present-day Edited American English*. <https://sketchengine.eu/brown-corpus/#toggle-id-1>
- Gervás, Pablo (2001). *Generating Poetry from a Prose Text: Creativity versus Faithfulness* [sesión de conferencia]. Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science. University of York.
- (2000). *WASP: Evaluation of Different Strategies for the Automatic Generation of Spanish Verse* [sesión de conferencia]. Proceedings of the AISB'00 Symposium on Creative & Cultural Aspects of AI & Cognitive Science. University of Birmingham.
- Goodwin, Jonathan y Holbo, John (Eds.). (2011). *Reading Graphs, Maps, and Trees: Responses to Franco Moretti*. Parlor Press LLC.

- Hart, M. S. (1971). *Proyecto Gutenberg*. <https://www.gutenberg.org/>
- Hernández-Lorenzo, Laura; De Sisto, Mirella; Pérez, Álvaro; De la Rosa, Javier; Ros, Salvador; y González-Blanco, Elena (2021). The Automatic Quantitative Metrical Analysis of Spanish Poetry with Rantanplan: A Preliminary Approach. *Tackling the Toolkit*, 31.
- Marchetti, Alessandro; Sprugnoli, Rachele; y Tonelli, Sara (2014). *Sentiment Analysis for the Humanities: The Case of Historical Texts*. Digital Humanities - DH, 2014.
- Martínez-Gamboa, Ricardo (2016). Big Data en Humanidades Digitales: de la escritura digital a la “lectura distante”. *Revista Chilena de Literatura*, 94, 39-58.
- (2015). Patrones cuantitativos en novelas chilenas de los siglos XIX a XXI. *Onomázein*, 32, 239-253.
- Moreno Sandoval, Antonio (1998). *Lingüística computacional*. Síntesis.
- Moretti, Franco (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.
- Navarro-Colorado, Borja (2017). A Metrical Scansion System for Fixed-metre Spanish Poetry. *Digital Scholarship in the Humanities*, 33(1), 112-127. <https://doi.org/10.1093/lc/fqx009>
- Navarro-Colorado, Borja; Ribes Lafoz, María; y Sánchez, Noelia (2016). *Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation* [sesión de conferencia]. Proceedings of the Language Resources and Evaluation Conference. http://www.lrec-conf.org/proceedings/lrec2016/pdf/453_Paper.pdf
- (2015). *Corpus of Spanish Golden-Age Sonnets*. Universidad de Alicante. <https://github.com/bncolorado/CorpusSonetosSigloDeOro>
- Pennebaker, James W.; Francis, Martha E.; y Booth, Roger J. (2001). *Linguistic Inquiry and Word Count: LIWC 2001*. Lawrence Erlbaum Associates.
- POSTDATA. (2020). *PoetryLab Corpus*. <https://mla.hcommons.org/deposits/item/hc:31763/>
- Real Academia Española (2014). *Corpus del español del siglo XXI (CORPES XXI)*. <https://www.rae.es/banco-de-datos/corpes-xxi>
- (2001). *Nuevo tesoro lexicográfico de la lengua española*. <https://apps.rae.es/ntlle/SrvltGUISalirNtile>
- (1994). *Corpus de referencia del español actual (CREA)*. <https://corpus.rae.es/creanet.html>
- ŠeĽa, Artjoms; Plecháč, Petr; y Lassche, Alie (2022). Semantics of European Poetry is Chaped by Conservative Forces: The Relationship between Poetic Meter and Meaning in Accentual-syllabic Verse. *PLOS One*, 17(4), e0266556. <https://doi.org/10.1371/journal.pone.0266556>
- Yu, Bel (2008). An Evaluation of Text Classification Methods for Literary Study. *Literary and Linguistic Computing*, 23(3), 327-343.

Ricardo Martínez-Gamboa, Valeria Araya Lopetegui, Daniel Mora, Nicolás Acosta y Suyli Aprosio



Esta obra está bajo licencia |internacional
Creative Commons Reconocimiento-NoComercial 4.0.