



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA - ESCUELA DE LETRAS - BIBLIO-  
TECAS UC

# **ANÁLISIS ESTILOMÉTRICO EN PYTHON APLICADO A EPISTOLARIOS DE GABRIELA MISTRAL**

**NATALY RUIZ CORNEJO**

Tesis presentada a la Dirección de Investigación y Postgrado  
como parte de los requisitos para optar al grado de  
Magister en Procesamiento y Gestión de la Información

Profesor Supervisor:  
**MAURICIO ARRIAGADA**

Santiago de Chile, Enero 2019

© MMXIX, NATALY RUIZ CORNEJO



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA - ESCUELA DE LETRAS - BIBLIO-  
TECAS UC

# **ANÁLISIS ESTILOMÉTRICO EN PYTHON APLICADO A EPISTOLARIOS DE GABRIELA MISTRAL**

**NATALY RUIZ CORNEJO**

Miembros del Comité:

MAURICIO ARRIAGADA

CÉSAR ANTONIO AGUILAR

JULIÁN SAN MIGUEL

Tesis presentada a la Dirección de Investigación y Postgrado  
como parte de los requisitos para optar al grado de  
Magister en Procesamiento y Gestión de la Información

Santiago de Chile, Enero 2019

© MMXIX, NATALY RUIZ CORNEJO

*Dedicado a usted lector que tiene  
un trozo de mi tiempo en sus manos.*

## **AGRADECIMIENTOS**

Agradezco de corazón a mi entorno; novio, familia y amigos, quienes me ayudaron, apoyaron y entendieron en este andar, créanme que la construcción de esta duna es gracias a cada grano de arena que me aportaron todos ustedes.

Agradezco al profesor Mauricio Arriagada por instarme y apoyarme en la realización de este trabajo.

Agradezco al Ministerio de las Culturas, las Artes y el Patrimonio, y al programa de Fondo del Libro y la Cultura por la oportunidad de realizar este magíster.

“No entiendo hasta hoy cómo el buen señor me abrió su tesoro, fiándome libros de buenas pastas y de papel fino”

Gabriela Mistral

## INDICE GENERAL

AGRADECIMIENTOS . . . . .	IV
INDICE DE FIGURAS . . . . .	VII
RESUMEN . . . . .	IX
ABSTRACT . . . . .	1
1. INTRODUCCIÓN . . . . .	2
1.1. Justificación . . . . .	3
1.2. Objetivo . . . . .	4
1.3. Hipótesis . . . . .	4
2. ESTADO DEL ARTE . . . . .	5
2.1. Contexto histórico . . . . .	7
2.1.1. Biografía de Gabriela Mistral . . . . .	8
2.1.2. Biografía de Manuel Magallanes Moure . . . . .	11
2.1.3. Biografía de Pedro Prado . . . . .	12
2.1.4. Grupo Los Diez . . . . .	14
2.1.5. Biografía de Doris Dana . . . . .	16
2.2. Epistolarios . . . . .	17
2.3. Búsqueda de Software para analizar los datos . . . . .	21
2.3.1. Corpus . . . . .	23
2.3.2. Word Cloud, Nube de palabras . . . . .	25
2.3.3. Estilometría . . . . .	27
2.3.4. Sketch Engine . . . . .	30
2.3.5. Phyton . . . . .	31
2.3.6. Natural Language Toolkit (NLTK) . . . . .	33
2.3.7. Curvas características de operación . . . . .	34

2.3.8. Chi square . . . . .	35
2.3.9. Método Delta . . . . .	36
3. DESARROLLO . . . . .	38
3.1. Transcripción de las cartas . . . . .	38
3.2. Estudio de los datos . . . . .	39
3.2.1. Corpus . . . . .	40
3.2.2. Trabajo en Python . . . . .	44
3.2.3. Aplicación del estudio . . . . .	45
4. ANÁLISIS DE RESULTADOS . . . . .	57
4.1. Word Cloud . . . . .	59
5. CONCLUSIÓN . . . . .	64
BIBLIOGRAFIA . . . . .	66

## INDICE DE FIGURAS

2.1. Nube de palabras, diseño común. . . . .	26
2.2. Nube de palabras, diseño personalizado. . . . .	27
2.3. Curvas características de operación. . . . .	35
3.1. Corpus Prado . . . . .	41
3.2. Corpus Magallanes Moure . . . . .	42
3.3. Corpus Doris Dana . . . . .	42
3.4. Conectores . . . . .	43
3.5. Set de datos del ejercicio. . . . .	45
3.6. 1er set, 1er test, Gráfico A, Pedro Prado. . . . .	46
3.7. 1er set, 1er test, Gráfico B, Manuel Magallanes. . . . .	47
3.8. 1er set, 1er test, Gráfico C, Doris Dana. . . . .	47
3.9. 1er set, 2do test, Gráfico A, Pedro Prado. . . . .	49
3.10. 1er set, 2do test, Gráfico B, Manuel Magallanes. . . . .	49
3.11. 1er set, 2do test, Gráfico C, Doris Dana. . . . .	50
3.12. 2do set, 1er test, Gráfico A, Pedro Prado. . . . .	51
3.13. 2do set, 1er test, Gráfico B, Manuel Magallanes. . . . .	52
3.14. 2do set, 1er test, Gráfico C, Doris Dana. . . . .	52
3.15. 2do set, 2do test, Gráfico A, Pedro Prado. . . . .	54
3.16. 2do set, 2do test, Gráfico B, Manuel Magallanes. . . . .	54
3.17. 2do set, 2do test, Gráfico C, Doris Dana. . . . .	55
4.1. Utilizando 1 muestra en el test case. . . . .	58
4.2. Utilizando 3 muestras en el test case. . . . .	59
4.3. Nube de palabras, Pedro Prado . . . . .	60

4.4. Nube de palabras, Manuel Magallanes . . . . .	61
4.5. Nube de palabras, Doris Dana . . . . .	62



## RESUMEN

El presente trabajo es un estudio estilométrico, que utiliza como data set una muestra conformada por epistolarios de la poetisa chilena y premio nobel Gabriela Mistral dirigidos a Pedro Prado, Manuel Magallanes Moure y Doris Dana. Este estudio es una réplica del ejercicio propuesto por François Dominic Laramée “Introducción a la estilometría con Python” en el cual se utilizan 3 formas de estudios estilométricos aplicados en Python.

En este estudio se valida la eficacia y eficiencia de Python para reconocer estilos y tendencias a través del estudio estilométrico, y se puede establecer y determinar la naturaleza de la relación que existía entre Mistral con Pedro Prado, Manuel Magallanes y Doris Dana.

Este trabajo es un input para la investigación de diferentes áreas del conocimiento, que tengan interés de continuar develando la figura e imaginario de Gabriela Mistral, a través de algún análisis automático utilizando herramientas de programación y ciencias de datos.

**Palabras Claves:** NLTK, estilometría, python, Gabriela Mistral, epistolarios, nube de palabras

## ABSTRACT

The following paper is a stylometric study, which uses as a data set a real sample of manuscripts of the Chilean poet and Nobel Prize winner Gabriela Mistral that make up correspondence to Pedro Prado, Manuel Magallanes Moure and Doris Dana. This study is a replica of the exercise proposed by François Dominic Laramée "Introduction to stylometry with Python" where 3 forms of stylometric studies applied in python are used.

This work validates the effectiveness and efficiency of Python to recognize styles and trends through stylometric study, and it can establish and determine the nature of the relationship that existed between Mistral with Pedro Prado, Manuel Magallanes and Doris Dana.

This paper aims to be an input for the investigation of different areas of knowledge, which have an interest in continuing to reveal the figure and imaginary of Gabriela Mistral, through an automatic analysis using programming tools and data science.

**Keywords:** NLTK, stylometry, python, Gabriela Mistral, correspondence, word cloud

## 1. INTRODUCCIÓN

El presente trabajo es un estudio estilométrico realizado en Python, utilizando biblioteca NLTK, basado en el ejercicio propuesto por François Dominic Laramée (Laramée, 2018), titulado “Introduction to stylometry with Python”, el cual fue publicado en la revista electrónica y abierta Programming Historian. Cabe destacar que todo lo utilizado en este trabajo fue en base a programas de código abierto.

Se seleccionó como objeto de estudio, 3 epistolarios pertenecientes a Gabriela Mistral, en los cuales se pueden estudiar sus expresiones afectivas y/o amorosas hacia otro ser. Las razones por las que se seleccionó a esta autora son las siguientes: su relevancia nacional e internacional como escritora, Mistral fue la primera mujer latinoamericana en ganar el Nobel de Literatura, fue un ejemplo en las letras para la escena nacional, fue una intelectual de avanzada respecto a su época, publicó sobre muchísimos temas vedados. A modo de ejemplo cabe señalar que fue una de las pocas las mujeres que trató y dio dignidad al tema de la maternidad en estado de soltería, poco publicado pero cotidiano en la sociedad chilena de entonces, y del hoy.

La figura de Mistral sigue teniendo un velo en la sociedad chilena; aún es reconocida en el inconsciente nacional por las rondas y canciones de cuna, por su mala suerte en el amor, por su imagen maternal y sus logros literarios. Sin embargo al estudiarla afloran su pasión por el amor, el interés en la educación, desde el aula y desde las políticas públicas y su interés por la figura y desarrollo de la mujer, entre tantos otros temas.

La data utilizada es una muestra de 3 epistolarios escritos por Mistral. Estos epistolarios fueron escritos a Manuel Magallanes Moure, Pedro Prado y Doris Dana, algunas cartas que conforman estos epistolarios son parte de la colección de Bibliotecas UC, fue una gran ayuda el tomarlas desde las publicaciones que se han realizado con ellas, ya que fue un ahorro de tiempo en la investigación y transcripción. Los destinatarios fueron seleccionados tomando en cuenta: la amistad de la poetisa con Pedro Prado, a quien comunica desde diferentes lugares y etapas de su vida, la relación de amor platónico con el poeta Magallanes Moure en el cual se manifiestan tópicos como el amor, pasión, culpa y resignación.

Por último se selecciona el epistolario dirigido a Doris Dana, figura tomada como su leal asistente, a quien, luego de ser develado el epistolario, se le dio el rol de amante por los más avezados y de asistente/amiga por los más recatados, y protectores de la biografía oficial.

Este trabajo es una muestra de la relación de las 3 áreas que conforman el Magíster de Procesamiento y Gestión de la Información de la Pontificia Universidad Católica, ya que trabaja y se vincula con las áreas de Ingeniería, aplicando herramientas de tecnologías de información, Letras (aplicando técnicas de procesamiento de texto como la estilometría), y Bibliotecología (ya que utiliza material de estudio perteneciente a las colecciones de Bibliotecas UC).

### **1.1. Justificación**

El presente trabajo y sus resultados, plantean 3 aportes que lo caracterizan para ser utilitario en diferentes aspectos:

- Entregar un modelo de uso de estilometría en epistolarios.
- Trabajar con una muestra real de manuscritos de un autor nacional.
- Ser el primer modelo estilométrico basado en epistolarios pertenecientes a Gabriela Mistral.

Este trabajo puede ser de interés para diferentes áreas del conocimiento, y un input para la investigación, desde distintos ángulos:

1. Bibliotecológico; por el trabajo investigativo que pueden fomentar las bibliotecas a través del material que resguardan. Y dando la posibilidad al mismo personal bibliotecario a colaborar en investigaciones interdisciplinarias utilizando sus propias colecciones como base.
2. Biográfico; por los aspectos psicológicos que el estudio puede revelar de la escritora y su uso del lenguaje epistolar cotidiano, además de relevar la influencia de relaciones y acontecimientos personales del universo mistraliano y además explorar el controvertido tema de identidad de género que envuelve a la Mistral.

3. Histórico; comprensión del contexto nacional y mundial en el cual se enmarca la producción de la obra, pudiendo identificar hitos y personajes relevantes. por la información relevante que se puede encontrar de la vida de la autora, y su vínculo con su trabajo de docente, política y escritora.
4. Literario; desde la aplicación de un punto de vista literario a un corpus ajeno al canon lírico de la autora. Por corpus ajeno nos referimos a los epistolarios de la autora que provienen desde lo íntimo, y que registran los cambios identitarios por la evolución de los signos y estados de ánimo plasmados en la escritura epistolar de Gabriela Mistral. Aplicación de técnicas de procesamiento de lenguaje natural, ejemplo a replicar en el área de la lingüística.

La investigación aporta a un enfoque de investigación basado en la aplicación de tecnologías de estudios estilométricos.

## **1.2. Objetivo**

1. Establecer y determinar, a través del estudio estilométrico, la naturaleza de la relación que existía entre Mistral con Pedro Prado, Manuel Magallanes Moure y Doris Dana utilizando herramientas de ciencias de datos y estilometría para contribuir al debate y la reconstrucción de la biografía e imaginario en torno a Gabriela Mistral.

## **1.3. Hipótesis**

1. El ejercicio propuesto por Laramée es replicable en epistolarios de Gabriela Mistral; dirigidos a Magallanes Moure (1914 a 1923), Pedro Prado (1915-1939), y Doris Dana (1948-1956).
2. Gracias a la aplicación de este análisis estilométrico se pueden detectar rasgos estilísticos asociados a dos tópicos literarios concretos: uno referido hacia el amor pasional (o Eros), y el otro hacia la amistad (el equivalente al concepto griego de "Philos": tener una veneración o aprecio a algo o alguien), en la literatura epistolar de Gabriela Mistral.

## 2. ESTADO DEL ARTE

Este trabajo es presentado en el marco de optar al Magíster en Procesamiento y Gestión de Información (*Magíster en Procesamiento y Gestión de Información*, s.f.), de la Pontificia Universidad Católica de Chile y tiene la característica de ser interdisciplinario ya que está desarrollado por la Facultad de Ingeniería, la Facultad de Letras y el Sistema de Bibliotecas UC. El presente trabajo es un reflejo de la relación de las 3 áreas ya que la idea inicial fue confeccionar algo utilizando material de Bibliotecas UC, con el objetivo de visibilizar y fomentar la investigación con material disponible en las colecciones de Bibliotecas UC, las cartas de Gabriela Mistral están ubicadas en la Biblioteca de Humanidades (*Biblioteca de Humanidades - Sistema de Bibliotecas UC*, s.f.), específicamente en la Sala Patrimonial, resguardadas bajo estrictos cuidados de humedad y temperatura para su correcta preservación, este material tiene la característica de ser único ya que la mayoría de ellas están escritas por puño y letra de la poetisa.

Para estudiar el material escrito, específicamente los epistolarios, se pensó en una manera fácil de realizar esto, ya que debía ser transcrito y luego analizado, por ende era necesario consultar en el área de Letras qué se podía hacer o cómo se podía realizar el estudio, la respuesta fue aplicar la estilometría, la cual se resume como una metodología estadística para analizar textos a través de la frecuencia del uso de palabras. José López en la publicación ¿Qué es la estilometría y para qué sirve? (López, 2012), indica que la Estilometría analiza ciertos rasgos del estilo del autor y los utiliza para comparar dos o más textos. El punto de base de la estilometría, según López, es que el estilo es algo que nace en el subconsciente, y por esta razón, cada quien tiene su estilo propio, se conoce que cada autor en su trabajo deja una huella lingüística que lo identifica y diferencia de otros autores, por lo cual cada estilo es único. La estilometría es una forma de analizar textos a diferentes niveles, algunas de sus aplicaciones son determinar la autoría de una obra, la autenticidad, clasificación de textos, medición de frecuencia de palabras e identificación de lenguas. El plus de la estilometría es el análisis del estilo de un autor a través de la recuperación de

datos estadísticos, por lo que aún diferentes áreas del conocimiento como son las letras y la estadística.

Un software reconocido en el área de letras para estudios estadísticos es R, sin embargo existe una tendencia y desarrollo de estudios estilométricos y reconocimiento de autorías utilizando Python en unión con la librería NLTK. Python es un lenguaje de programación que tiene un uso generalizado hoy en día en el área de las Ciencias de la Computación.

Respecto a la característica del material de análisis, el trabajo con epistolarios es un aporte para estudiar al autor desde su espacio privado ya que hay palabras, trazos y conexiones léxicas que van más allá de su obra y que plasma cuando se comunica, en un epistolario podemos ver la huella del estilo del autor.

Middleton (Murry, 1956), indica que “lo valioso de un estilo personal dependerá de que sea o no la expresión de un auténtico sentimiento individual” y que lo que parezca artificial en la obra del autor será un triunfo del arte, entendiendo la artificialidad como el trabajo del autor quien, a pesar de que utilice un lenguaje natural para expresarse, lo hace de una manera original, permitiéndole comunicar sentimientos y crear mundos a través de sus herramientas únicas e individuales rozando el arte y la creación, a diferencia del concepto artificial que tiene la intención de destacar pero con el anhelo egocéntrico de alcanzar la perfección. Existe una delgada línea entre estos artificios ya que, indica Middleton Murry “Todo estilo es artificial en el sentido de que todos los buenos estilos se logran por artificio”, aún así destacan los buenos estilos cuando se logra la naturalidad en la creación, ya que se entrelaza lo intelectual con lo emotivo.

La poesía es conocida como un medio escrito para expresar las emociones personales, y una carta es un escrito que una persona dirige a otra, para comunicar y/o expresarse.

Un epistolario es un conjunto de cartas y en él se pueden ver aspectos desconocidos del autor, respecto a su vida, relaciones y pensamientos. Un ejemplo de esto, ligado a los 3 epistolarios estudiados es que al revisarlos se puede ver el desarrollo de la identidad de Gabriela Mistral y cómo se va asentando su personalidad e identidad con el pseudónimo escogido. Además de lo novedoso que es para los estudiosos de la poetisa, estudiar la

evolución de la identidad de género que se expresa en Mistral a lo largo de sus epístolas y cómo podría cambiar la biografía oficial de ésta.

Un ejemplo del aporte que genera la revisión de un epistolario es que en la biografía de Gabriela Mistral se entrega el dato de que ésta realizó conferencias en La Habana y en Panamá, y al leer sus epístolas se puede inferir los lugares visitados y su opinión sobre ellos, por ejemplo algo tan banal como lo es el clima del lugar.

Este trabajo es un aporte al estudiar a Gabriela Mistral desde un ángulo relativamente nuevo, el cual comenzó oficialmente con los estudios realizados a la muestra tangible del epistolario dirigido a Doris Dana, aquellos tratan de la homosexualidad de Mistral. Este estudio aporta con una mirada estadística a aquellos estudios entregando una visión objetiva (datos), respecto a la relación de Mistral con Doris Dana. El resultado del trabajo puede tomarse para reconstruir su figura apoyándose de la tecnología, para comprender a la Mistral como figura en el ambiente literario nacional, es más, fue el primer chileno en ganar el Nobel de Literatura, dato relevante para una mujer que por su género no tuvo la relevancia merecida en su país natal. También sirve para rescatar y destacar su figura de reformista, de maestra, su figura como ente político, y su legado a la identidad chilena, ya que su patrimonio es un aporte a la historia de Chile.

## **2.1. Contexto histórico**

A continuación se presentan las biografías de manera orgánica acorde a la historia y los sucesos que relacionan a los poetas en cuestión, dado que el eje central de esta investigación es Gabriela Mistral, quien conoce a Pedro Prado y este se relaciona, a través del Grupo de Los Diez con Magallanes Moure, quien cierra el círculo con Mistral, cabe señalar que la base de las biografías fueron tomadas del sitio Memoria Chilena. (*Personajes - Memoria Chilena, Biblioteca Nacional de Chile, s.f.*).



### 2.1.1. Biografía de Gabriela Mistral

Gabriela Mistral es el seudónimo de la poetisa Lucila de María del Perpetuo Socorro Godoy Alcayaga, nació el 7 de abril de 1889 en el pueblo de Vicuña, Región de Coquimbo, específicamente Valle del Elqui, lugar que se compone de pequeños pueblecitos, que serán determinantes en esta figura, y en su inspiración.

Nació en el seno de una familia conservadora, más por la madre, que por el padre quien las abandonó cuando Lucila contaba con 3 años de edad. Su madre, Petronila Alcayaga, se caracterizaba por un carácter sedentario y estable. La familia de Lucila contaba de 3 integrantes: ella, su madre y su media hermana, 11 años mayor, de nombre Emelina. Ya que era una familia pequeña fueron muy unidas, Emelina ejerció gran influencia en Lucila respecto a su interés por la educación y los niños, y su madre en las letras. Al momento de fallecer su madre, Lucila le dedicó la primera sección de su libro *Tala* (Mistral, 2013), a la cual nombró “Muerte de mi Madre”:

“Amados pechos que me nutrieron con una leche más que otra viva... ¡resucidad, resucidad, si existe la hora, si es cierto el día, para que Cristo os reconozca y a otro país deis alegría...”

Mistral recibió sus primeras lecciones en casa, con su hermana Emelina, y más tarde en la escuelita de La Unión. Su madre la envió a Vicuña para terminar la enseñanza básica, en donde realizó la labor de lazarillo de la directora, misma mujer que le diagnosticó “retraso mental” y provocó en contra de Lucila escenas escolares de enorme humillación, este factor fue determinante para iniciar su formación autodidacta.

En 1904 Lucila colaboró en el periódico *El Coquimbo de La Serena*, utilizando diferentes seudónimos, como “Alguien”, “Soledad” y “Alma”, para sus primeros poemas publicados como: *Ensoñaciones*, *Carta Íntima* y *Junto al Mar*. Los títulos elegidos para sus poemas, así como también los seudónimos con que firmó, expresaron el carácter solitario, introvertido y romántico de la joven poetisa. A los 15 años (1904), fue nombrada ayudante en la Escuela de La Compañía, cerca de Vicuña. En 1908 ya trabajaba como maestra en

la escuelita de La Cantera. Allí conoció a Romeo Ureta, un empleado de Ferrocarriles, conocido como uno de los grandes amores de Lucila Godoy. Por lo demás Gabriela Mistral, para seguir su trayectoria en la educación, quiso ingresar a la Escuela Normal de Preceptoras de La Serena, lo cual se vio frustrado debido a la resistencia que despertaron algunos poemas suyos en círculos conservadores locales, que los calificaron como "paganos" "socialistas". En 1910 se trasladó a Santiago, donde trabajó en la Escuela de Barranca luego de haber aprobado los exámenes especiales en la Escuela Normal de Preceptores. A partir de ese momento empezó a trabajar en distintas escuelas alrededor del país, como las de las ciudades de Traiguén, Punta Arenas, Antofagasta y Temuco.

En 1912 se trasladó al Liceo de Antofagasta, donde enseñó Historia y además fue inspectora general. Ese mismo año fue nombrada inspectora y profesora de Castellano en el Liceo de Los Andes.

Su primer gran éxito literario fuera del ámbito regional: ocurrió el 12 de diciembre de 1914, cuando obtuvo la más alta distinción en los Juegos Florales de Santiago con sus Sonetos de la Muerte, cabe destacar que este momento marcó un punto en su relación epistolar con Magallanes Moure quien era jurado del certamen.

En 1918 fue nombrada profesora de Castellano y directora del Liceo de Punta Arenas.

Se cree que el seudónimo de Gabriela Mistral fue inspirado por el poeta italiano Gabriele D'Annunzio, a quien admiraba y tomó de nombre. Y tal como aparece en el sitio de Memoria Chilena sobre el apellido Mistral, existen diferentes versiones que lo explican: una, por el poeta Federico Mistral; y otra versión señala que Mistral proviene del nombre de un viento provenzal. Pero la expresión literaria de Gabriela Mistral no dejó dudas acerca de sus concepciones románticas y religiosas, a la vez que de su gran capacidad imaginativa.

En 1922 fue invitada a México por el Ministerio de Educación de ese país, con el fin de participar en los planes de la reforma educacional mexicana, y en la organización y fundación de bibliotecas populares. Mistral destaca el buen recibimiento en ese país, y por lo demás, hasta el día de hoy destaca que ese país la haya llamado a trabajar y aportar

en su política de educación, sin haber sido llamada en Chile a pesar de ya ser conocida internacionalmente.

En 1923 publicó *Lecturas para Mujeres*, en México. La Editorial Cervantes de Barcelona publicó una antología titulada "Las Mejores Poesías".

De regreso al país la Universidad de Chile decidió otorgar a Gabriela Mistral el título de profesora de Castellano (1923).

En 1924 se embarcó por primera vez a Europa, publicó en Madrid *Ternura*, pequeño volumen de versos dedicados a los niños.

Al año siguiente estaba de vuelta en Latinoamérica y recorrió Brasil, Uruguay y Argentina. De regreso a Chile, se le otorgó una pensión y jubiló como maestra.

En 1926 participó como representante de Chile en el Instituto de Cooperación Intelectual de la Liga de las Naciones en Ginebra. Fue nombrada Secretaria de una de las Secciones Americanas de esa misma organización. También estuvo en Suiza en un Congreso de Educadores en Lucarno (1927).

En 1930 viajó a Estados Unidos donde llevó a cabo actividades pedagógicas, como seminarios para la secundaria. Un año después recorrió algunas naciones centroamericanas donde continuó con cátedras sobre literatura. En la Universidad de Puerto Rico desarrolló una charla sobre literatura hispanoamericana. Además realizó otras conferencias en La Habana y en Panamá (lugares que nombra entre sus epistolarios).

En 1932, Gabriela Mistral fue designada Cónsul particular de libre elección, trasladándose a Génova. Una vez que declaró su posición antifascista, dejó su cargo. Pero en 1933 fue incorporada al mismo en Madrid y después en Lisboa.

Por su importante cooperación en actividades culturales, y su gran desempeño consular, en 1935, bajo el gobierno de Alessandri, se le designó Cónsul de elección con carácter vitalicio, por Ley del Congreso del 4 de septiembre de ese año.

En forma paralela a su carrera consular, trabajaba como colaboradora en *El Mercurio* de Santiago, *Crítica* de Buenos Aires, *El Tiempo* de Bogotá, *El Universal* de Caracas, y en

el Puerto Rico Ilustrado. En 1936 viajó a Oporto y después a Guatemala como encargada de negocios y cónsul general.

La recaudación que se obtuvo por la publicación de Tala en Buenos Aires (1938) fue destinada a instituciones que albergaron a niños españoles afectados por la Guerra Civil española.

En Brasil, específicamente el 14 de agosto de 1943, Juan Miguel, apodado Yin Yin y amado hijo de alma de Gabriela Mistral, quien era realmente su sobrino pero a quien ella crió como un hijo al quedar huérfano, se suicidó, tenía 17 años en aquel entonces y su decisión dejó a Gabriela elevando plegarias en su nombre, con el alma rota y en depresión.

El 15 de noviembre de 1945, Mistral se convirtió en la primera poetisa y literata hispanoamericana y chilena en ganar el Premio Nobel de Literatura, de hecho fue la primera mujer chilena en ganar el Nobel y la primera ciudadana chilena en recibirlo ya que 26 años después lo haría Neruda. Tras 6 años del reconocimiento internacional llegó el aletargado reconocimiento nacional, en 1951 Chile le concedió el Premio Nacional de Literatura y 5 años después en 1956, se realizó un homenaje oficial a su figura y trabajo.

Lucila Godoy falleció de cáncer el 10 de enero de 1957 en el Hospital General de Hampstead, en Nueva York. Sus restos fueron traídos a Chile el 19 de enero y descansan en el Valle del Elqui, en Montegrande, Región de Coquimbo, en donde ella añoraba volver, como anécdota Mistral llevaba un cofre con un puñado de tierra del Valle, la que se ha convertido en un souvenir del lugar.

### **2.1.2. Biografía de Manuel Magallanes Moure**

Nació el 8 de noviembre de 1878, en La Serena (Región de Coquimbo, Chile). Su padre fue Intendente de Coquimbo y miembro de la Corte de Apelaciones. Estudió en Santiago y vivió largos años en la ciudad de San Bernardo, en la calle Eyzaguirre en donde realizaba tertulias literarias.

Se casó en 1903 con Amalia Villa Magallanes, quien era su prima. En su vida laboral destaca su labor como editor de la revista Chile Ilustrado (1902), su participación en la

*Revista Cómica*, su colaboración en *El Mercurio*, *Las Ultimas Noticias* y la *Revista Zigzag*. Además se desempeñó como Secretario Municipal y Alcalde de San Bernardo. Fundó en 1911 el periódico *La Reforma* que se mantuvo hasta 1916. Fue Presidente de la Asociación de Artistas y Escritores de Chile la que en 1914 patrocinó Los Juegos Florales, donde Gabriela Mistral presentó sus *Sonetos de la Muerte* y ganase la más alta distinción, tema que se habla en el epistolario entre él y Mistral.

Magallanes Moure participó en el Grupo Literario Los Diez junto a Pedro Prado y los otros artistas nombrados. Incursionó en la pintura, presentando una exposición pública en 1916. En 1918 funda la Sociedad de Conferencias Públicas en el salón de la Escuela Superior de Niñas N°2. En 1922 viajó en misión oficial a Europa.

La poesía de Manuel Magallanes Moure se enmarca dentro del contexto del romanticismo, lejos del realismo criollo y del naturalismo crudo de los escritores de su época, se caracteriza por sus versos sencillos, simples, de mucho sentimiento, donde el paisaje, especialmente todo lo concerniente al mar (buques, playas, olas), cobra especial realce. Su trabajo el cual Gabriela admiraba y su belleza fue el conjunto que la atrajo.

Manuel Magallanes Moure murió a los 45 años, en Santiago de Chile, en la misma comuna y casa donde se estableció, San Bernardo, en la calle Eyzaguirre en 1924.

### **2.1.3. Biografía de Pedro Prado**

El poeta Pedro Prado nació en Santiago el 8 de octubre de 1886. Estudió en el Instituto Nacional e ingresó a la escuela de Arquitectura, sin culminar sus estudios. A los 20 años, cuando falleció su padre, quien era médico y agricultor, empezó a ejercer alternadamente sus tres profesiones sin título, o más bien sus oficios: arquitecto, agricultor y escritor. Su poesía surgió en un contexto donde interactúan dos poéticas antagónicas; por una parte, el modernismo rubendariano, de espíritu cosmopolita y referentes franceses, y por otra, un nacionalismo que se tradujo, en la lírica chilena, en la exaltación de los valores regionales.

El libro *Flores de cardo*, publicado hacia el 1908, produjo en la crítica una mezcla de admiración y estupor, pues se acababa de introducir en Chile, con él, el culto al verso libre y la ruptura de las sujeciones métricas, lo que significó el punto de partida de muchas y sustanciales transformaciones en la lírica nacional.

Pedro Prado pertenecía a la Generación Literaria de 1920, promoción de poetas que comienza a publicar sus primeras obras entre 1907 y 1917, junto a Ernesto Guzmán, Carlos R. Mondaca, Manuel Magallanes Moure, Max Jara y Juan Guzmán Cruchaga, grupo de poetas más bien intimistas, volcados a la vida interior o paisajes provincianos y campesinos, sin mayores desbordes vanguardistas ni experimentación formal.

Pedro Prado, fundó en 1915, el grupo “Los Diez”, famoso grupo intelectual de aquella época, que con el objetivo de publicar, crearon la revista “Los Diez”, revista de tan solo 4 números.

Su obra *Pájaros errantes*, publicada en 1915, es considerada por muchos críticos como la obra poética cumbre de Pedro Prado. Acorde a lo que se infiere en las cartas y la apreciación de Ana María Cuneo (Cuneo, 2017); la obra de Prado es una de las favoritas de Mistral lo que la llevó a hablarle desde la esquina de una admiradora de su prosa, y tratarlo como *maestro*, Mistral le escribe:

"Los jóvenes que escriben hoy en Chile ignoran que tienen en usted ya un maestro i sería la mayor de mis satisfacciones aprender en usted sobriedad, alta originalidad, profundidad i cuántas cosas más"(junio 1915).

La obra estrictamente poética de Pedro Prado se suspendió en 1915, con la publicación de *Los pájaros errantes*, para reanudarse diecinueve años después, en 1934, con los sonetos *Camino de las horas*, la cual le valió el Premio Academia de Roma de la Embajada de Italia en Chile, luego de esto volvió a publicar en 1945, obra titulada *Esta bella ciudad envenenada*. Durante este periodo, Prado produjo numerosas y clásicas obras en prosa, desde ensayos hasta novelas, de las que sobresalen *Alsino*, vasta novela alegórica y poética sobre el tema de la libertad del individuo, *La reina rapanui*, y la novela *Un juez rural*, texto de corte realista y con toques de un humor singular.

Pedro Prado fue el 8vo personaje en recibir el Premio Nacional de Literatura, en 1949. Murió víctima de un síncope en Viña del Mar el 31 de enero de 1952.

#### **2.1.4. Grupo Los Diez**

Según las palabras de Pedro Prado el origen de Los Diez fue mediante una conversación con Julio Bertrand Vidal, quien asombrado por la alegría y el entusiasmo del poeta, le preguntó si existían más hombres como él. Prado, entonces sonrió y contestó que tal vez podrían encontrarse unos diez más. Ante esta respuesta, Bertrand contestó: "Me gustaría conocerlos...". el grupo lo integraban pintores, escultores, músicos, arquitectos y poetas.

Estos 10 hombres, que en la práctica fueron más que diez, a los que hacía referencia Pedro Prado eran: Manuel Magallanes Moure (poeta, cuentista, pintor), Juan Francisco González (pintor), Armando Donoso (crítico literario, periodista), Alberto García Guerrero (músico), Alberto Ried (poeta, cuentista, escultor, pintor), Acario Cotapos (músico), Augusto D'halmar (novelista, cuentista), Alfonso Leng (músico y compositor), Julio Ortiz de Zárate (pintor) Ernesto A. Guzmán (poeta y ensayista), Eduardo Barrios (novelista y dramaturgo) y Julio Bertrand Vidal (arquitecto y pintor).

La primera sesión oficial de Los Diez, se llevó a cabo en la Biblioteca Nacional, ocasión en la cual Pedro Prado, dio lectura a la Somera iniciación al Jelsé. En este texto, clara y presumiblemente un manifiesto, Prado señaló: "Los Diez no forman ni una secta, ni una institución, ni una sociedad. Carecen de disposiciones establecidas, y no pretenden otra cosa que cultivar el arte con una libertad natural". Además, estableció una suerte de perfil, que en cierta forma resumiría la actitud de sus integrantes frente al arte y frente a la vida, se indica "Es requisito imprescindible para pertenecer a Los Diez estar convencidos de que nosotros no encarnamos la esperanza del mundo; pero (...), debemos observar con prolijidad todo nuevo ser que se cruce en nuestro camino, por si él encarnase esa esperanza, lo que no impide que, después de ese examen, él y nosotros nos riamos con gran pesadumbre y bulliciosa algarazara de los continuos engaños que por este motivo nos ocurran".

Debido al carácter multidisciplinario de esta agrupación, sus obras alcanzaron casi todo el espectro del arte. Así, se hizo posible notar su presencia en la literatura, pintura, música, arquitectura, política, e incluso llegaron a diseñar su propio calendario y revista (2016-2017), toda esta producción fue el resultado de sus tertulias.

En agosto de 1917, tras la publicación de *Pobrecitas* de Armando Moock, Los Diez terminaron su labor editorial. En una carta dirigida a Nathanael Yáñez Silva a propósito de un artículo en el que se aludía a la supuesta muerte del grupo, Manuel Magallanes Moure, señaló: "Los Diez no son una institución formada más o menos artificialmente, ni una sociedad cuyos miembros estén amarrados por algún nudo reglamentario, de esos que una vez que se cortan o que se desatan, producen el desparramamiento (...) Nuestra unión tiene una más firme atadura: nos unen el arte y la amistad. No tenemos obligaciones que llenar ni compromisos que cumplir; nos acerca el placer de estar juntos".

La revista *Los Diez* fue el resultado editorial que la reunión amistosa de un grupo de artistas e intelectuales (artistas plásticos, artistas visuales, músicos y poetas), tuvo durante las primeras décadas del siglo XX, todos con la intención de fomentar la creación y difusión del arte propiamente chileno. De este importante cenáculo de la cultura nacional surgieron compositores, intérpretes y gestores culturales que animaron posteriormente la vida musical del país durante aquella época.

La *Revista Los Diez* llevada al papel en la Imprenta Universitaria, arrojó su primer número al público en septiembre 1916 y el último en abril de 1917. Su propósito era publicar 12 volúmenes al año pero alcanzó a editar sólo 4 números alternados con 8 libros (o ediciones especiales).

Gabriela Mistral era una ávida lectora de la *Revista Los Diez*, y mantenía lazos con varios participantes de la agrupación, un ejemplo de esto es lo que ella indica en un extracto de una carta dirigida a Pedro Prado (Mistral, 1993):

"Distinguido amigo, no he recibido ni la carta a la que usted alude en su cordial de reciente fecha ni ENSAYOS, libro suyo del cual sólo he sabido por la nómina de obras que viene en LOS DIEZ."



Con una periodicidad de un número al mes, el deseo de los artistas mancomunados en esta causa era, como la aclara su editorial, convertirse "(...) en un portavoz completo, serio y digno de todos los que en Chile se dedican, por imperiosa necesidad de espíritu y con nobleza artística, a producir obras de calidad"(Castro, 1965).

Al igual que otras revistas de la época, el tono de la revista no era sólo de difusión sino que también de crítica y opinión, con una impronta nacionalista abierta a los artistas "latinoamericanos y extranjeros".

Estructuralmente, la publicación poseía ocho secciones en diverso orden. Primero una editorial (idéntica en cada número), luego una sección de poemas, otras de cuentos y una de partituras de compositores nacionales. Más adelante estaba la sección de artículos de análisis, teoría literaria y arte en general, para luego dar paso a la crónica, la crítica artística y una sección de asuntos variados.

Lo interesante de la revista *Los Diez* era la ambición poética de su contenido. Aparentemente desvinculado de la realidad, la revista se comunicaba con la sociedad civil por medio de una prosa llena de un sentimiento desgarrador y romántico, con un hálito de alejamiento del mundo y en una constante intención de autonomía artística.

#### **2.1.5. Biografía de Doris Dana**

Doris Dana, nació en el seno de una familia aristocrática de la sociedad de Nueva York en 1920, su madre fue Alberta Webster, canadiense, y su padre William Sheperd Dana, quienes tuvieron 3 hijas donde Dana era la del medio.

Doris Dana estudió en el Lennox School, y la secundaria en el exclusivo Bryn Mawr College, conocido como uno de los centros de estudios para mujeres más elitistas del país. Obtuvo un grado de Bachiller en Literatura en el Barnard College de la Universidad de Columbia.

Posteriormente fue la primera profesional de su género en redactar un programa de reeducación para la Alemania de posguerra en 1945.

Dana fue benefactora de Gabriela Mistral desde 1946 hasta el día de su muerte en 1957. Posteriormente por testamento, Dana fue receptora, representante y recelosa albacea de la obra y documentos de Gabriela Mistral hasta su fallecimiento en 2006, adquiriendo la representación su sobrina, Doris Atkinson quien donó gran parte del material a Chile.

## 2.2. Epistolarios

La Real Academia Española (Real Academia Española, 2019), indica que un epistolario es un “Libro o cuaderno en que se hallan recogidas varias cartas o epístolas de un autor o de varios, escritas a diferentes personas sobre diversas materias”, en este trabajo se seleccionaron muestras de cartas pertenecientes a 3 epistolarios escritos por Gabriela Mistral a 3 destinatarios.

En el libro *En batalla de sencillez* (Mistral, 1993), explican su significado desde una visión más personal, dice que el “género epistolar consiste en un supuesto diálogo en donde las personas se turnan para monologar” las personas al confeccionar una epístola tienen la clara intención de expresarse y de dirigirse a un otro, es un flujo comunicativo plasmado en lo impreso, en el texto se indica que la “mutua adecuación” de los seres humanos para contactarse y conectarse con un otro, esperando su tiempo para responder o para leer y adecuando una actitud formal o de cercanía según sea la ocasión y el destinatario, esa capacidad humana es una expresión de belleza y estética plasmada en un epistolario, más aún si tomamos en cuenta que al estudiarlos el investigador tiene acceso a una expresión inconsciente del autor, una fotografía de su estilo, pensamiento y redacción.

Los autores de la recopilación de *En batalla de sencillez* (Mistral, 1993), comparten características destacables del epistolario dirigido a Pedro Prado, en donde se recopilan 24 años de epístolas (1915 a 1939), lo describen como un epistolario “cortés, decoroso, afable, entusiasta y noble”.

El epistolario de Mistral dedicado a Manuel Magallanes Moure tiene una anécdota biográfica y de pasión: mientras ambos se escribían, mantenían una relación platónica y Mistral le profesaba su amor, Magallanes Moure estaba casado, lo cual le producía culpa a

Mistral por el tinte pecaminoso de la acción y el sentir. Este epistolario comenzó en 1913 y se consolidó luego que Mistral ganase el Juego de las Flores, en donde Magallanes fuese uno de los jurados. El compromiso de los autores de este epistolario era destruir las cartas recibidas, Gabriela Mistral cumplió lo dicho mientras que Manuel las guardó, finalmente la familia de Magallanes, en manos de Amalia Redondo Magallanes (nieta) compartió este material: 38 misivas que permitieron la primera publicación de este epistolario en manos de Sergio Fernández Larraín: *Cartas de amor, de Gabriela Mistral* (Mistral, 1978) y *Cartas de amor y desamor* (Mistral, 1999), además de la publicación de María Ester Martínez y Luis Vargas Saavedra del 2005 titulada *Manuel, en los labios por mucho tiempo* (Mistral, 2005). En cuanto a las otras misivas inéditas, 24 fueron adquiridas por el Instituto de Letras de la Universidad Católica de Chile (CELICH), que actualmente son parte del acervo de Bibliotecas UC, y se encuentran en las estanterías de la Colección Patrimonial de Biblioteca de Humanidades en el Campus San Joaquín, y las últimas 11 forman parte de la colección de la Universidad de Notre Dame (Estados Unidos).

Respecto al tercer epistolario, la donación de las cartas de Doris Atkinson a Chile se concretó en 2007, como indica EMOL (*Doris Atkinson concreta traspaso a Chile de material de Gabriela Mistral*, 2007), en noviembre del 2007, “A las 14:30 horas fue firmado en la Embajada de Chile en Washington el documento que acredita la cesión de todo el material mistraliano que se encontraba en Estados Unidos”, esta donación incluía: libros, primeras ediciones, manuscritos, cartas, fotografías y distintos materiales, algunos de estos se encuentran exhibidos en el Museo de Gabriela Mistral en Vicuña ya que gran parte fue donada a la Dirección de Bibliotecas, Archivos y Museos, y la Orden Franciscana. Donde se estipuló que los libros existentes en su residencia de USA fuesen parte de la colección del Museo. La segunda, en tanto, administrará los derechos y regalías de su obra, en beneficio de los niños pobres de Montegrando, además de recibir un broche de plata y turquesa que Mistral usó para recibir el Premio Nobel, tres biblias con su firma y dos crucifijos que había conservado Doris Dana, y que serían expuestos en el Museo Franciscano de Santiago. El embajador Mariano Fernández agradeció en nombre del gobierno: "Hoy es un día muy

importante para Chile, por el regreso del gran legado de Gabriela Mistral a nuestro país, gracias a la generosidad y sabiduría de Doris Atkinson".

Las cartas en sí son considerados objetos valiosos porque algunos son manuscritos de puño y letra de Mistral, lo cual los hace únicos, los otros son mecanografiados y el contenido de estos es una fuente interminable de material para conocer más a la Mistral y la posibilidad de investigar desde diferentes ángulos. Una opinión que resume el contenido de estos epistolarios es la del especialista Luis Vargas Saavedra a EMOL (Rodríguez, 2013), quien indica que en las cartas:

"Se evidencia la creciente seguridad en sí misma como persona y poeta, más una inteligencia emotiva y expresiva: el don de poder decir cabal y hermosamente lo que se siente. En suma, el genio inexplicable de una provinciana sin educación completa"

En las cartas podemos encontrar conversaciones con temáticas de: pensamientos, viajes, trabajo, opiniones, proyectos, arrebatos, sentimientos, situación literaria de la autora, y entorno y contexto del ambiente literario de aquellos tiempos, entre otros temas. Además a través de los lugares de destino de las cartas podemos ver la ubicación de la Mistral y sus opiniones de los lugares en los que se encuentra.

Los contenidos de las cartas levantaron el velo de la orientación sexual de Mistral, el cual nunca fue revelado por ella o por su entorno, es por esto que los epistolarios se tomaron como un objeto de estudio, escrutando su contenido y relacionándolo con su vida, identificando las palabras y los destinatarios, para corroborar o desestimar las suposiciones respecto a sus preferencias sexuales y sacando el tema a flote, ya que en biografías de ella, anteriores al siglo XXI no se tocaba el tema de sus relaciones amorosas o su sexualidad o género, ya que se utilizaba o trabajaba la figura de una Mistral heterosexual, con mala suerte en el tema amoroso, poeta sin darle mucho hincapié a la Mistral reformista o política, entre otras. Como un ejemplo del trabajo de la figura de la mistral y el velo dado al tema de sus relaciones amorosas en las publicaciones o estudios biográficos, anteriores, es que por muchos años se consideró a Romelio Ureta como su gran amor, otorgándole un tinte trágico y novelesco a la historia.

La imagen de la Mistral ha sido manipulada durante muchos años dando al imaginario colectivo ciertas características según la época, situaciones tratadas en el texto *A Queer Mother for the Nation* (Fiol-Matta, 2002), lo que levantó un sin número de artículos y publicaciones, como la del New York Times en 2003 (Rohter, 2013), donde además se escribió abiertamente sobre su lesbianismo, temática que repercutió luego en la prensa en español, por ejemplo en el diario El tiempo del mismo año, con un artículo titulado *Gabriela Mistral: poeta y lesbiana* (Gabriela Mistral: poeta y lesbiana, 2003) o en la BBC con el titular *La prefieren loca que lesbiana* (Mundo, 2016).

Al estudiar las cartas se puede observar el desarrollo de la firma de la Mistral, como una señal de identidad; comenzó firmando como L., o Lucila y terminó firmando como Gabriela, pasando por Lucila Godoy, GabrMistral, GMistral y Gabriela Mistral.

Además de ver la evolución de la firma de Gabriela Mistral desde el punto de vista de la identidad como autora, también se puede observar en las cartas la identidad de género de la Mistral, la cual se comenzó a discutir desde la década de los 90 y oficialmente a comienzos del siglo XXI con la donación de Atkinson y el acceso al material epistolario, ya que queda plasmada en las cartas, donde se diferencia el uso del sujeto femenino para referirse a ella misma en las cartas dedicadas a Magallanes Moure entre el 1915-1923 y las dirigidas a Doris Dana en los últimos años de su vida, en donde se identifica como un sujeto masculino, como ejemplo podemos tomar las frases: “Te beso, tuyo” y “¡Qué estúpido ha sido el que más te quiere, Doris mía!”.

Eugenia Brito (Brito, 2014) comenta que “este tránsito de lo femenino a lo masculino es una de las características del epistolario e indica un punto de fuga de la escritura mistraliana” ya que en sus publicaciones siempre produjo desde el autor femenino, tocando y trabajando temáticas que en aquellos tiempos eran tierras netamente femeninas.

La obra de la Mistral y sus acervos personales han generado cuantiosas investigaciones y publicaciones en formato electrónico e impreso, ya que se pueden estudiar diversos temas ligados al perfil de esta; ya sea biografías, publicaciones sobre identidad de género, literatura infantil, del área poética, su rol político, educacional, entre otros. Sin embargo

los epistolarios dieron una fuente más de publicaciones y develaron otra Gabriela como la mujer que expresaba su pasión hacia Magallanes Moure que dio la posibilidad de publicar y conocerla desde ese ámbito recién a finales de la década de los 70 y esa Gabriela que expresa amor a Dana recién la pudimos conocer en la década del 2000. Los contenidos de los epistolarios han dado pie a investigaciones plasmadas en artículos, libros, tesis, presentaciones, entrevistas, entre otros. Sin embargo, no se ha realizado un estudio estilométrico de ellas, lo que confiere a este trabajo la oportunidad de ser el primero en su categoría.

Por lo que contiene y lo que conlleva el estudio de los epistolarios es que se considera que son un objeto indispensable de estudiar ya que son la fuente viva e inconsciente de la autora que nos revela su esencia, y por lo demás este material es posible de estudiar objetivamente a través de la comunión de las letras y la estadística bajo la estilometría.

### **2.3. Búsqueda de Software para analizar los datos**

Se buscó definiciones de las palabras y temas a tratar en el trabajo, además de diferentes programas y fórmulas, a continuación se describen las temáticas más relevantes que se decidió utilizar en la confección y desarrollo del presente estudio:

1. Corpus
2. Word Cloud
3. Estilometría
4. Sketch Engine
5. Phyton
6. Natural Language Toolkit (NLTK)
7. Curvas características de operación
8. Chi square
9. Método Delta

Las herramientas de análisis textual permiten el tratamiento automático de la información para apoyar el estudio empírico de las lenguas. Permiten la consulta rápida de una o

varias colecciones de textos electrónicos, así como la preparación y el mantenimiento de bases de datos textuales. Los textos pueden analizarse desde dos niveles de complejidad:

1. El nivel de datos, es el nivel más simple desde el punto de vista operacional. El objeto de análisis es el texto puro, entendido como un conjunto de caracteres. Las operaciones automáticas que pueden realizarse son, únicamente, aquellas basadas en la manipulación de los caracteres del texto. Por ejemplo, la localización de todas las palabras que comienzan por un prefijo determinado, las palabras que acompañan a otra dada (colocaciones), o la frecuencia de aparición de un término.
2. El nivel de información se relaciona con la interpretación de los textos. A este nivel corresponden las operaciones que necesitan disponer, además de los textos, de la interpretación de los mismos, como por ejemplo la localización de todos los verbos transitivos en una colección de textos o los textos que tratan sobre “las enfermedades del salmón”. Este tipo de consultas más “inteligentes” requieren que el texto tenga marcado de alguna forma este tipo de información. Es importante tener en cuenta que el procesamiento automático de la información es más complicado que el procesamiento de datos y que sólo puede realizarse si previamente se ha preparado (pre-procesamiento) el texto (datos) insertando la interpretación (semántica) de estos datos textuales. Uno de los mecanismos de pre-procesamiento es la inserción en el texto de marcas o etiquetas explícitas con la información asociada a cada elemento textual (proceso de etiquetado).

Todas las herramientas de análisis textual permiten el análisis de texto puro (datos), por ejemplo R. Otras como Sketch Engine, ofrecen la posibilidad de analizar, además, el texto marcado (información), otra herramienta muy utilizada y fácil de aprender es NLTK de Python.

Para trabajar con herramientas de análisis textual es necesario contar con una colección de textos representativos de aquello que se desea analizar. Esta colección representativa de textos se denomina corpus.

### **2.3.1. Corpus**

Biber (Biber, 1993), indica que los Corpus se conocen como grandes colecciones de textos aleatoriamente seleccionados de entre los textos de cierto género, tema, etc. Otra definición de corpus es la de la RAE (Real Academia Española, 2019) “Conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”. Se indica que una de las aplicaciones principales de los corpus es proporcionar el conocimiento léxico de diferentes tipos y niveles. Hay dos tipos de probabilidades a estudiar en los corpus, según indican Gelbukh, Sidorov y Chanona (Gelbukh, s.f.):

- Probabilidades absolutas: ¿cuál es la probabilidad de la situación X en el texto? Por ejemplo, ¿cuál es la frecuencia relativa de la palabra ecuación en el texto?
- Probabilidades condicionales: ¿cuál es la probabilidad de la situación X en la presencia de la situación Y? Es decir, la investigación estadística se aplica a un sub-corpus que consiste de los contextos que contienen la situación Y. Por ejemplo, ¿cuál es la probabilidad de la palabra ecuación dado que la siguiente palabra es diferencial?

Para un estudio empírico bien fundamentado y unos resultados significativos, es necesario que la colección sea lo suficientemente grande y equilibrada como para abarcar la mayor cantidad de ocurrencias posibles del fenómeno a estudiar. En el documento del Taller sobre herramientas de análisis textual: La herramienta Sketch Engine (Facultad de Filología, 2011), se pueden distinguir varias tipologías de corpus.



Se puede establecer una tipología de corpus en función de los criterios utilizados para la clasificación. Estos criterios pueden ser de diferente índole:

1. La modalidad de la lengua.
2. La naturaleza física de los datos.
3. La cobertura y la temática.
4. El número de lenguas.
5. La disponibilidad y el modo de recopilación.

1. Según la modalidad de la lengua:

- Corpus de lengua escrita.
- Corpus de lengua hablada. Transcripciones ortográficas de grabaciones, utilizadas sobre todo en lingüística de corpus. Grabaciones y transcripciones ortográficas y/o fonéticas, empleadas en fonética y tecnologías del habla.
- Corpus mixtos.

2. Según la naturaleza física de los datos:

- Corpus textuales o escritos: muestras de lengua escrita.
- Corpus orales: Transcripciones ortográficas de grabaciones, Grabaciones y transcripciones ortográficas y/o fonéticas.
- Corpus multimodales (imagen y texto).

3. Según la cobertura y la temática:

- Por períodos
- Por géneros
- Por temas
- Por autores
- Por registros

4. Según el número de lenguas:

- Corpus monolingües: están formados por textos de una sola lengua. Se recopilan con el objetivo de dar cuenta de una lengua o variedad lingüística.

- Corpus bilingües o multilingües: están formados por textos de dos (bilingües) o más (multilingües) lenguas sin que, en principio, sean traducciones unos de otros y sin compartir criterios de selección.
  - Corpus comparables (“paired texts”): consisten en una selección de textos en más de una lengua o variedad lingüística parecidos en cuanto a sus características y que comparten criterios de selección. Se utilizan sobre todo para comparar variedades de la lengua en estudios contrastivos.
  - Corpus paralelos (“bi-texts”): recogen textos en más de una lengua (bilingües o multilingües) pero, a diferencia de los anteriores, se trata del mismo texto traducido a una o más lenguas.
5. Según la disponibilidad y el modo de recopilación:
- Corpus disponibles como recursos en instituciones o en grupos de investigación.
  - Corpus contruidos a partir de textos seleccionados y recopilados manualmente según criterios concretos y para estudiar unos aspectos concretos.
  - Corpus recopilados automáticamente por herramientas que permiten una búsqueda y recopilación a través de internet mediante el uso de palabras semillas.

### **2.3.2. Word Cloud, Nube de palabras**

Yuping Jin (Jin, 2017) define word cloud o nube de palabras como “un tipo de lista ponderada para visualizar datos de idioma o texto”. Es considerada una rama importante dentro de la minería de datos por su aplicabilidad en variadas áreas. La nube de palabras cuenta con diversos e incluso personalizados modelos y se usa generalmente como mapas gráficos para visibilizar la frecuencia de uso de términos bajo un tamaño relativo de los mismos.

Los 3 algoritmos más utilizados para ser representados en nubes de palabras son:

- Las figuras 2.1 y 2.2 son ejemplos de la página GitHub (Mueller, 2018), de visualizaciones de nubes de palabras hechas con el algoritmo de frecuencia:

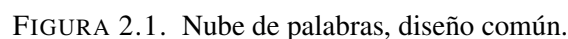




FIGURA 2.2. Nube de palabras, diseño personalizado.

La nube de palabras al ser visual se convierte en una herramienta amigable del big data para facilitar la comprensión de los resultados.

### 2.3.3. Estilometría

François Dominic Laramée (Laramée, 2018), define a la estilometría como el estudio cuantitativo del estilo literario a través de métodos computacionales de lectura a distancia. Se basa en la observación de que los autores tienden a escribir de manera relativamente consistente, reconocible y única. La estilometría es una técnica moderna del siglo XXI, que analiza ciertos rasgos del estilo de un autor. El punto

de base de la estilometría es que el estilo es inconsciente y deja una especie de huellas estilográficas, y por esta razón, cada quien tiene su estilo propio, tal como indica Eugenia Brito, “las cartas o las notas son parte de un proceso idéntico de escritura”. Algunas de sus aplicaciones es determinar la autoría de una obra, la autenticidad, clasificación de textos, medición de frecuencia de palabras, identificación de lenguas.

En el post de José López (López, 2012), se describen dos usos de la estilometría que son relativos a este trabajo: determinar la autoría y clasificar textos. Se entiende que cada texto tiene marcadores de estilo que hacen diferente o similar a cada texto a nivel de estilo. Los marcadores léxicos se dividen en dos:

- Riqueza del vocabulario
- Frecuencia de las palabras de función

Las palabras de función, por ejemplo, podrían ser simples preposiciones (a, contra, bajo, etc.). La idea es que la forma y el lugar donde se emplean dichas palabras es suficiente para identificar a alguien como el autor de una obra. Como resultado, con este tipo de análisis podría determinarse el plagio de una obra.

Asimismo, la estilometría sirve para clasificar textos en diferentes categorías. Por ejemplo, tipo de caracteres de un texto, idioma, etc.

Para este tipo de análisis, es común la utilización de n-grams, también denominados unidad de palabras múltiples o MWU de su sigla en inglés, es una secuencia de números de elementos (números, dígitos, palabras, letras, etc.). En el contexto de los corpus, los n-grams se referirán a secuencias de palabras. Por ejemplo:

- Unigrama = una palabra
- Bigrama = Secuencia de dos palabras
- Trigrama = secuencia de tres palabras, etc.
- N-grama = Secuencia de N palabras, según se estime conveniente.

Los elementos dentro de un n-gram pueden no tener ninguna relación entre ellos aparte del hecho de que aparecen uno al lado del otro, a diferencia de las *colocaciones* que son palabras con una relación entre ellas pero que no necesariamente están ubicadas de manera anexa, de ser así podrían considerarse un ngram.

El estudio de los n-grams es importante para la traducción automática ya que su frecuencia indica frases correctas en comparación con palabras aisladas, también son un aporte en el aprendizaje de idiomas.

Finalmente, una vez obtenidos los datos, n-grams, se almacenan con el fin de analizar los datos obtenidos.

Algunos parámetros de la muestra que se deben tomar en cuenta para que el análisis estilométrico tenga mayor fiabilidad son:

- Cantidad de textos.
- Cantidad de autores.
- Extensión de n-grams

Para realizar este tipo de análisis estadístico, existen varias herramientas que pueden facilitar la obtención de los datos, una de las más destacadas y utilizadas en la literatura es el análisis estilométrico en R, este programa está bajo licencia GNU (Licencia Pública General, en pocas palabras es un software libre y de código abierto), que proporciona varias técnicas de análisis lineal y no lineal, así como representación gráfica. Otro software es Sketchengine, producto pagado que pertenece a la compañía Lexical Computing, empresa creada en 2003 y dedicada al estudio del lenguaje a través de la lingüística computacional.

Miguel ángel Bravo (Bravo, 2018), citando a Stamatatos explica que la estilometría ha demostrado que las palabras gramaticales, conocidas como palabras de función, o palabras de función pequeña; como las preposiciones, los artículos, conjunciones, entre otros, tienen un mayor peso de estilo a la hora de establecer una huella

lingüística (o estilística), de una obra o autor en comparación a los términos semánticos, debido a que las palabras de función, independientemente del género textual o de la temática, se usan inconscientemente.

Este trabajo será una ejemplo de la utilización de un software relativamente nuevo, Sketchengine, para medir la frecuencia y cantidad de palabras, cada corpus se considerará un texto, además este trabajo servirá para aplicar 3 enfoques estilométricos utilizando Python.

#### **2.3.4. Sketch Engine**

Sketch Engine (lexical computing, s.f.), es la herramienta de análisis de texto en línea, que funciona con grandes muestras de lenguaje (corpus de texto), es utilizada para explorar cómo funciona el lenguaje a través de algoritmos que analizan textos auténticos de miles de millones de palabras para identificar el lenguaje o términos utilizados. También está diseñada para el análisis de texto o aplicaciones de minería de textos.

En la página de Sketch Engine se indica que contiene 500 corpus para usar en más de 90 idiomas, cada uno con un tamaño de hasta 30 mil millones de palabras para proporcionar una muestra de idioma verdaderamente representativa. Sus funciones son:

- Generar listas de palabras
- Extracción de palabras claves y términos
- Terminología bilingüe
- Calcular N-GRAMS
- Identificar neologismos
- Etiquetado de texto
- Creación de tesauros

Las características de esta herramienta son: funcionalidades derivadas, escalabilidad, soporte multilingüe y la capacidad de trabajar con grandes los corpus y que

estos sean en varios idiomas, da la posibilidad de buscar de diferentes maneras; desde la simple hasta la lematizada y por CQL (Contextual Query Language).

### 2.3.5. Phyton

Python (*Python 3.7.2 documentation*, 2019), es un lenguaje de programación muy utilizado a nivel internacional, es aplicable en muchos ámbitos y es fácil de aprender. Cuenta con estructuras de datos eficientes y de alto nivel y un enfoque simple pero efectivo a la programación orientada a objetos. La elegante sintaxis de Python y su tipado dinámico, junto con su naturaleza interpretada, hacen de éste un lenguaje ideal para scripting y desarrollo rápido de aplicaciones en diversas áreas y sobre la mayoría de las plataformas. Este lenguaje de programación fue creado por el holandés Guido van Rossum a finales de los 80 (1989 específicamente), y debe su nombre al grupo de humoristas Monty Python (si usted alguna vez creyó que era en base al nombre de la temida serpiente, entonces somos  $n+1$ ). La creación de van Rossum responde a las siguientes características / necesidades:

- Python debería ser fácil, intuitivo y tan potente como sus principales competidores.
- El proyecto sería de código abierto para que cualquiera pudiera colaborar.
- El código escrito en Python sería tan comprensible como cualquier texto en inglés.
- Python debería ser apto para las actividades diarias permitiendo la construcción de prototipos en poco tiempo.

Python cuenta con el plus de ser código abierto, en el tutorial de Python v.2.7.0 se indica que el intérprete de Python y la extensa biblioteca estándar están a libre disposición en forma binaria y de código fuente para las principales plataformas desde



el sitio web de Python, además de estar disponible documentación muy completa la cual puede ser utilizada como herramienta de aprendizaje .

El intérprete de Python puede extenderse fácilmente con nuevas funcionalidades y tipos de datos implementados en C o C++ (u otros lenguajes accesibles desde C). Python también puede usarse como un lenguaje de extensiones para aplicaciones personalizables, por último, cabe destacar que sus actualizaciones siempre intentan ser compatibles con versiones anteriores .

En la página de lenguaje de programacion.net (*Guía de aprendizaje de Python. Programación en Castellano.*, 2019), se destacan las siguientes características de Python:

- La sintaxis de Python es muy fácil de usar, de ahí que se ha ido haciendo cada vez más popular en los años recientes.
- Hay lenguajes que siguen siendo muy populares, como Java o C++, pero Python se está usando mucho más en varios campos, y sobretodo está ganando popularidad para la programación de páginas html, gracias a Django.

En el tutorial oficial de Python en español (Rossum, 2017), se señala lo siguiente:

- Lenguaje de propósito general: Se logra crear páginas sin la necesidad de tener un alto conocimiento del lenguaje de programación.
- Es multiparadigma: su fuerte es la programación orientada a objetos, sin embargo también funciona como programación imperativa (con sentencias de bucle) o programación funcional (con módulos y funciones).
- Es un lenguaje interpretado: Cuando programamos en Python, no compilamos el código fuente a código máquina, sino que hay un intérprete que es el que ejecutará el programa basándose en el código directamente. Esto facilita que sea multiplataforma.

- Es multiplataforma: Python se puede usar en distintos sistemas operativos, ya que se han creado intérpretes para Unix, Linux, Windows y sistemas Mac Os.
- Es de tipado dinámico: Las variables se adaptan al ejecutar el programa, no es necesario indicar tipo de datos (int, string, float, etc.).
- Es orientado a objetos: se puede aplicar otro estilo de programación, reutilizar los componentes gracias a la herencia y sus funciones de polimorfismo.
- Su comunidad: Dispone de un gran soporte gracias a su comunidad, quienes comparten tutoriales para aprender desde 0 o resolver dudas.

Programas conocidos mundialmente, hechos con Python:

- Instagram
- Youtube
- Pinterest
- Dropbox

El presente trabajo se basó en la publicación de François Dominic Laramée “Introducción a la estilometría con Python” (*Natural Language Toolkit — NLTK 3.4 documentation*, 2019), la clase fue publicada en abril del 2018, y tiene el propósito de enseñar de manera online y gratuita a realizar un análisis estilométrico utilizando Python y la biblioteca NLTK, para determinar autoría de textos en disputa. Para esto se utiliza el compilado El Federalista, el cual tiene autoría en disputa y se entrega el data set de los documentos para que el usuario pueda trabajar y practicar el ejercicio, el cual propone 3 métodos para aplicar el contenido: las curvas características de composición de Mendenhall, el método Chi-cuadrado de Kilgariff y el método Delta de John Burrows.

### **2.3.6. Natural Language Toolkit (NLTK)**

Según la página oficial de NLTK (*Natural Language Toolkit — NLTK 3.4 documentation*, 2019), se indica que es la plataforma líder para la creación de programas Python para trabajar con datos en lenguaje humano. Proporciona interfaces

fáciles de usar para más de 50 recursos corporales y léxicos, además de un conjunto de bibliotecas de procesamiento de texto para clasificación, tokenización, derivación, etiquetado, análisis y razonamiento semántico. Esta plataforma se encuentra disponible para Windows, Mac OS X y Linux. Lo mejor de todo es que NLTK es un proyecto gratuito, de código abierto y dirigido a la comunidad.

NLTK es considerada una herramienta muy apta para enseñar y trabajar en el área de la lingüística computacional utilizando Python” y “una biblioteca poderosa para trabajar con lenguaje natural”.

François Laramée (Laramée, 2018), indica que Nltk funciona bien siempre que un idioma proporcione una forma clara de distinguir los límites de las palabras dentro de una palabra, sin embargo en idiomas como el chino para los que no existe una distinción clara entre los límites de las palabras, y son más conceptuales, el uso de Nltk puede ser problemático.

Sus características más relevantes es que es una herramienta fácil de utilizar, rápida de aprender y fácil de montar ya que evita tener que desarrollar una gran cantidad de infraestructura de software, es por esto que es considerado un módulo muy popular en Python aplicado al procesamiento de lenguaje natural.

### **2.3.7. Curvas características de operación**

En el ejercicio propuesto por Laramée se utilizó esta gráfica y este ejercicio en honor al trabajo realizado por Thomas Corwin Mendenhall, quien fue uno de los primeros en realizar un análisis cuantitativo de distribución de frecuencia de palabras, por ende uno de los primeros estudios estilométricos, el cual fue realizado en el siglo XIX y a mano. En el ejercicio se explica que si se cuenta la longitud de palabras utilizadas en variados segmentos que tuviesen entre 1,000 o 5,000 palabras de cualquier texto, y luego se traza un gráfico de las distribuciones de longitud de palabras, las curvas se verían más o menos iguales sin importar qué partes o segmentos del texto se hubiesen seleccionado.

Más conocidas como CO, son ampliamente utilizadas en Control Estadístico de Procesos (CEP), ya que la curva característica de operación es una muestra que presenta la potencia de discriminación de un plan de muestreo de aceptación (Carro Paz, s.f.). La curva grafica las probabilidades de aceptar un lote (elemento inspeccionado), versus la fracción defectuosa del mismo lote. Para el caso del ejercicio el vector izquierdo del gráfico representa las longitudes de palabras frecuentes. Las curvas características de operación grafican las probabilidades de una muestra, y son visualmente fáciles de comprender, es por esto que Laramée las escogió para representar el trabajo y método de Mendenhall, retribuyendo el valor histórico de haber realizado un intento de estudio estilométrico hace ya más de un siglo y con la característica, inverosímil en nuestros tiempos, de ser manual. Esta técnica fue utilizada en el primer test de técnicas de atribución de autoría del ejercicio realizado por Laramée y replicado en este trabajo.

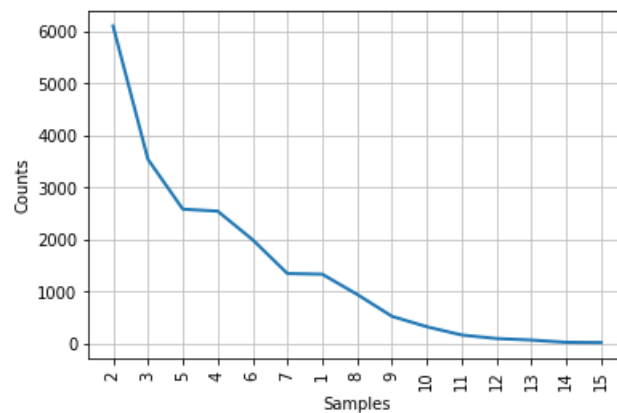


FIGURA 2.3. Curvas características de operación.

### 2.3.8. Chi square

Según lo expuesto en Credo Reference (*Chi-Square Test* | *The Concise Corsini Encyclopedia of Psychology and Behavioral Science* - Credo Reference, s.f.):

Chi Square, también representada por “ $\chi^2$ ” es una prueba que fue desarrollada por Karl Pearson en 1900, es considerada como uno de los avances más importantes en la historia de la estadística. La prueba y la distribución estadística en la que se basa

tienen una amplia variedad de aplicaciones en la investigación más humanista. Sus dos usos principales son; probar la independencia de dos variables y evaluar qué tan bien, un modelo teórico o un conjunto de probabilidades, a priori, se ajustan a un conjunto de datos. En ambos casos, la prueba de chi-cuadrado generalmente se considera un procedimiento no paramétrico que involucra las frecuencias observadas (O) y esperadas (E). Las frecuencias esperadas se pueden determinar teórica o empíricamente. La fórmula básica para calcular la estadística de chi-cuadrado es: La distribución de  $\chi^2$  está relacionada con la distribución normal, de modo que el cuadrado de una desviación normal estándar ( $z^2$ ) se distribuye como un  $\chi^2$  con un grado de libertad. La distribución chi-cuadrada también describe la distribución muestral de la varianza,  $s^2$ , tal que  $\chi^2 = (N - 1) s^2 / \sigma^2$  con  $N - 1$  grados de libertad.

### **2.3.9. Método Delta**

El texto de Laramée describe el método de Delta de la siguiente manera (Laramée, 2018): “Al igual que el Chi-cuadrado de Kilgariff, el Delta de Burrows es una medida de la distancia entre un texto cuya autoría queremos establecer y algún otro corpus”

El Método Delta está diseñado para comparar un texto anónimo (o un conjunto de textos), con las firmas de muchos autores al mismo tiempo. También otorga el mismo peso a cada característica que mide, evitando así el problema de las palabras comunes que generan ruido en los resultados, lo cual es un problema con las pruebas de Chi cuadrado. Estas dos características hacen del método Delta de John Burrows una solución efectiva al estudiar el problema de la autoría.

Como se indica en la publicación <sup>En</sup> “entendiendo Delta desde las humanidades” (Calvo Tello, 2016) Delta fue propuesta por John Burrows en 2002 para utilizarse en casos de atribución de autoría, presentando resultados exitosos, se basa en que la variación de frecuencia de las palabras más frecuentes en un texto permite reconocer autoría.

El funcionamiento del algoritmo de Burrows se puede resumir de la siguiente manera, para comprender los resultados que arroje el ejercicio:

- Encuentra las  $n$  palabras más frecuentes en el corpus para usar como características.
- Calcula la proporción de cada una de los subcorpus (cada una de esas  $n$  características,) de los autores representadas por cada una de esas  $n$  características, como un porcentaje del número total de palabras. Ejemplo, la palabra "the" puede representar el 4.72
- Calcula la media y la desviación estándar de estos valores de  $x$  y las usa como la media oficial y la desviación estándar para esta característica en todo el corpus.
- Calcula un puntaje  $z$ , para cada una de las  $n$  características y  $x$  subcorpus: esto identifica la distancia de la norma del corpus con el uso de una característica en un subcorpus en particular.
- Calcula los mismos puntajes  $z$  para cada característica en el texto que se está estudiando.
- Calcula una puntuación delta comparando el documento anónimo con el subcorpus de cada candidato. Esto se consigue al tomar el promedio de los valores absolutos de las diferencias entre las puntuaciones  $z$  para cada característica entre el papel anónimo y el subcorpus del candidato. Esto le da igual peso a cada función, sin importar la frecuencia con que aparezcan las palabras en los textos.

El resultado se observa en el puntaje delta entre el subcorpus del autor y el caso de prueba, el resultado más bajo, más cerca del 0, será quien se quede con la autoría.

### **3. DESARROLLO**

#### **3.1. Transcripción de las cartas**

Para esta investigación se utilizó una muestra de 84 cartas escritas o mecanografiadas por Gabriela Mistral. Parte de la muestra de cartas que se utilizaron en esta investigación fueron recopiladas desde la colección que posee Biblioteca de Humanidades de la Pontificia Universidad Católica de Chile, estas cartas están dirigidas a dos destinatarios:

Epistolario de Manuel Magallanes Moure (37 cartas), dirigido al poeta entre los años 1915 y 1923. Las primeras 24 cartas fueron transcritas por Amalia Redondo Magallanes, nieta de Magallanes Moure.

17 cartas dirigidas al escritor Pedro Prado Calvo, ganador del Premio Nacional de Literatura de 1949.

Ya que 20 estaban transcritas, se trabajaron las 34 restantes, se transcribieron con apoyo de la bibliotecóloga Javiera Bravo, quien posee conocimientos sobre transcripciones y letras manuscritas, especialmente las de Gabriela Mistral. Posteriormente se corroboró el trabajo con el libro *Manuel en los labios* por mucho tiempo y con el texto *En Batalla de sencillez* de Lucila a Gabriela.

El tercer epistolario es una muestra de 30 cartas dirigidas a Doris Dana que se recopilaron del texto *Niña Errante: cartas a Doris Dana*. Su transcripción no se realizó a mano, ya que se utilizaron los siguientes recursos detallados paso a paso:

- Digitalización de las páginas del libro a PDF, 300 ppp.
- Se cambiaron a formato word utilizando el software Acrobat reader.
- Cambio de formato a txt.

Finalmente cuando todas las cartas se transformaron a formato txt se les agregó el dato de fecha y lugar desde donde fueron escritas con el fin de rescatar datos georeferenciales con el objetivo de confeccionar un trabajo posterior, el orden de estos datos se plasmó en el siguiente orden: Lugar, año, mes

El criterio de ubicar el mes al último fue debido a que lugar y año siempre estaban presentes, el mes solo en algunas cartas y así se podía obtener la información 0 / 1 / 2 (opcional).

### **3.2. Estudio de los datos**

Se utilizó el software Sketchengine para probar una herramienta nueva a disposición, y por su facilidad de trabajar los datos para analizar los listados de palabras en los corpus.

El análisis estilométrico se realizará basándose en el ejercicio propuesto por François Dominic Laramée, titulado “Introduction to stylometry with Python” (Laramée, 2018). La decisión de trabajar con estas herramientas la expresa claramente el artículo escrito por Loper y Bird en “NLTK: The Natural Language Toolkit” (Loper, 2002):

- El lenguaje tiene una curva de aprendizaje superficial, de modo que los programadores novatos obtengan recompensas inmediatas por sus esfuerzos.
- El lenguaje debe ser compatible con la creación rápida de prototipos y un ciclo corto de desarrollo.
- El código debe ser autodocumentado, con una sintaxis y una semántica transparentes.
- Fácil escribir programas estructurados, idealmente orientados a objetos pero sin la carga asociada con los idiomas como C ++.
- El lenguaje debe tener una biblioteca de gráficos fácil de usar para admitir el desarrollo de interfaces gráficas de usuario.

Como lenguaje interpretado, Python es adecuado para prototipado rápido. El código de Python es excepcionalmente legible, y ha sido elogiado como "pseudocódigo ejecutable". Python es un lenguaje orientado a objetos, pero no de manera punitiva, y es fácil de encapsular datos y métodos dentro de las clases de Python

En el ejercicio propuesto por Laramée, se trabajan por etapas, 3 enfoques estilométricos diferentes:



- Curvas características de composición de Mendenhall
- Método Chi-cuadrado de Kilgariff
- Método Delta de John Burrows

Los materiales requeridos:

1. Conjuntos de datos: El conjunto de datos utilizado en este trabajo fueron los corpus extraídos de los epistolarios de Gabriela Mistral dirigidos a Pedro Prado, Manuel Magallanes Moure y Doris Dana.

Para aplicar el ejercicio estilométrico es necesario preparar y adaptar los archivos a estructuras de datos, se asignó un número correlativo a todas las cartas del corpus, este número se asigna al nombre del archivo en formato txt y dentro del documento, a modo de ejemplificar lo realizado: m\_1.txt hasta culminar todas las cartas m\_84.txt.

2. Software Anaconda, Jupyter y Python: En este ejercicio, al igual que la lección en el cual se basó, se utilizó la siguiente versión y biblioteca de idiomas de Python.
  - Python 3.6.4
  - Natural Language Toolkit (NLTK), spanish.
  - Matplotlib
  - 3 corpus o set de datos (conjunto de datos), el cual más adelante será denominado como Directorio.
3. Biblioteca NLTK: Biblioteca con contenido de idiomas para el procesamiento del lenguaje natural.

### 3.2.1. Corpus

Respecto al conjunto de datos, o corpus es necesario saber la siguiente información:

Características generales:

- De lengua escrita
- Textual

- Por autor
- Bilingüe
- Construido a partir de textos seleccionados

Características específicas:

- Número de textos: 84 documentos, entre manuscritos y textos mecanografiados
- Cantidad de autores: 1 autor

El corpus se dividió en 3 grupos, o 3 subcorpus, según destinatario:

1. Dirigidas a Pedro Prado, Figura (3.1)

- Idioma Español
- 17 cartas
- Palabras totales 5.521

GENERAL INFO		COUNTS		COMMON TAGS		LEMPOS SUFFIXES	
Language	Spanish	Tokens	6,570	noun	N.*	noun	-n
Tagset	<a href="#">Description</a>	Words	5,521	verb	V.*	verb	-v
Word sketch grammar	<a href="#">Description</a>	Sentences	324	adjective	A.*	adjective	-j
		Documents	17	adverb	R.*	adverb	-r
				pronoun	P.*	pronoun	-p

FIGURA 3.1. Corpus Prado

2. Dirigidas a Manuel Magallanes Moure, Figura (3.2)

- Idioma Español
- 37 cartas
- Palabras totales: 21.530

MISTRALMAGALLANES

user/nruizc/mistralmagallanos

MANAGE CORPUS

Cartas de Mistral a Magallanos

GENERAL INFO

Language

Spanish

Tagset

[Description](#)

Word sketch grammar

[Description](#)

COUNTS

Tokens

25,561

Words

21,530

Sentences

1,443

Documents

37

COMMON TAGS

noun

N.\*

verb

V.\*

adjective

A.\*

adverb

R.\*

pronoun

P.\*

LEMPOS SUFFIXES

noun

-n

verb

-v

adjective

-j

adverb

-r

pronoun

-p

FIGURA 3.2. Corpus Magallanes Moure

### 3. Dirigidas a Doris Dana, Figura (4.3)

- Idioma español e inglés
- 30 cartas
- Palabras totales: 12.114

MISTRALDANA

user/nruizc/mistraldana

MANAGE CORPUS

Cartas de Mistral a Doris Dana

GENERAL INFO

Language	Spanish
Tagset	<a href="#">Description</a>
Word sketch grammar	<a href="#">Description</a>

COUNTS

Tokens	14,460
Words	12,114
Sentences	1,052
Documents	30

COMMON TAGS

noun	N *
verb	V *
adjective	A *
adverb	R *
pronoun	P *

LEMPOS SUFFIXES

noun	-n
verb	-v
adjective	-j
adverb	-r
pronoun	-p

FIGURA 3.3. Corpus Doris Dana

- Términos totales de la sumatoria de los corpus: 39.165
- Cantidad de términos (conectores), a estudiar: 29.

Para la 3a prueba de John Burrows se estudiarán los conectores o palabras de función (3.4) propuestos en el ejercicio de Laramée traducidos al español:

de
que
la
no
a
en
y
el
me
yo
lo
un
es
te
por
una
los
i
mi
para
se
con
pero
las
tu
como
su
si
ud

FIGURA 3.4. Conectores

En la documentación del ejercicio se explica que: “Las formas en que los escritores usan palabras de función pequeña, como artículos, preposiciones y conjunciones, han resultado particularmente reveladoras. En un estudio de los métodos estilométricos históricos y actuales, Efstathios Stamatatos señala que las palabras funcionales son utilizadas por los autores en gran parte inconscientes, y son independientes del tema. Para el análisis estilométrico esto es muy ventajoso, ya que es probable que el patrón de inconsciencia varíe menos, sobre el cuerpo de un autor, que su vocabulario general.

### 3.2.2. Trabajo en Python

Este estudio se basó en la prueba realizada por François Dominic Laramée en su publicación “Introducción a la estilometría con Python”, el cual fue publicado en el 2018 y tiene dificultad media. El objetivo de este, a grandes rasgos, es:

Usar El Federalista, el cual es una colección de 85 artículos seminales de teoría política publicados entre octubre de 1787 y mayo de 1788, es conocido por ser la contribución más grande y duradera de Estados Unidos al campo de la filosofía política, el cual fue publicado de forma anónima, pero escrito por 3 autores, quienes después de mucho tiempo se adjudicaron la autoría de alguno de ellos, en donde hay autorías ambiguas que son el eje de este estudio de caso, que utilizando 3 enfoques estilométricos diferentes intenta develar. Los 3 enfoques se denominaron:

- Curvas características de operación de Mendenhall.
- Método Chi-cuadrado de Kilgariff.
- Método Delta de John Burrows.

Para desarrollar el estudio se dividieron los artículos de El Federalista en seis categorías:

1. Los 51 artículos que se sabe que fueron escritos por Alexander Hamilton.
2. Los 14 artículos que se sabe que fueron escritos por James Madison.
3. Cuatro de los cinco documentos que se sabe que fueron escritos por John Jay.
4. Tres documentos que probablemente fueron coescritos por Madison y Hamilton y de los cuales Madison reclamó la autoría principal.
5. Los 12 papeles disputados entre Hamilton y Madison.
6. Federalista 64 en una categoría propia.

Se aplicó código dependiente de idioma para dividir un texto en un conjunto de palabras en español; spanish.

Para el estudio de este trabajo, basándose en el estudio anterior se marcaron los 84 textos de manera correlativa para trabajar las cartas como strings (cadena de caracteres), como se aprecia en la figura (3.5).

- Prado 1-17
- Magallanes 18-54
- Dana 55-84

```
cartas = {  
    'Prado': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17],  
    'Magallanes': [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45],  
    'Dana': [55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 81, 82, 83, 84],  
    'test case': [79]  
}
```

FIGURA 3.5. Set de datos del ejercicio.

Test case, se denominará aleatoriamente según las necesidades de estudio.

Este directorio fue denominado data\_nata el cual contiene un tipo de datos compuesto por un número arbitrario de pares clave-valor; en este caso, los nombres de los autores servirán como claves, mientras que las listas de números de los documentos serán los valores asociados con estas claves.

- Primera Prueba de estilometría: Utilización de curvas características de operación de Mendenhall.
- Segunda prueba estilométrica: Kilgariff's Chi-Squared Method.
- Tercera prueba estilométrica: John Burrows' Delta Method (avanzada): esta prueba revisa la distancia entre un texto y los corpus.

La base para este trabajo fue identificar los textos como corpus y los test según los set de datos a realizar.

### 3.2.3. Aplicación del estudio

Las pruebas se dividieron y asignaron de la siguiente forma:

1er set de estudio:

- 1er test: Carta 79 escrita a Dana

- 2do test: Carta 39 escrita a Magallanes

2do set de estudio:

- 1er test: Cartas 18, 25 y 38 escritas a Magallanes
- 2do test: Cartas 57, 63 y 77 escritas a Dana

**A continuación se entregan los resultados de la aplicación de los test:**

1er set de estudio:

- 1er test: Carta 79 escrita a Dana
  1. Primera Prueba de estilometría: Utilización de curvas características de operación de Mendenhall.

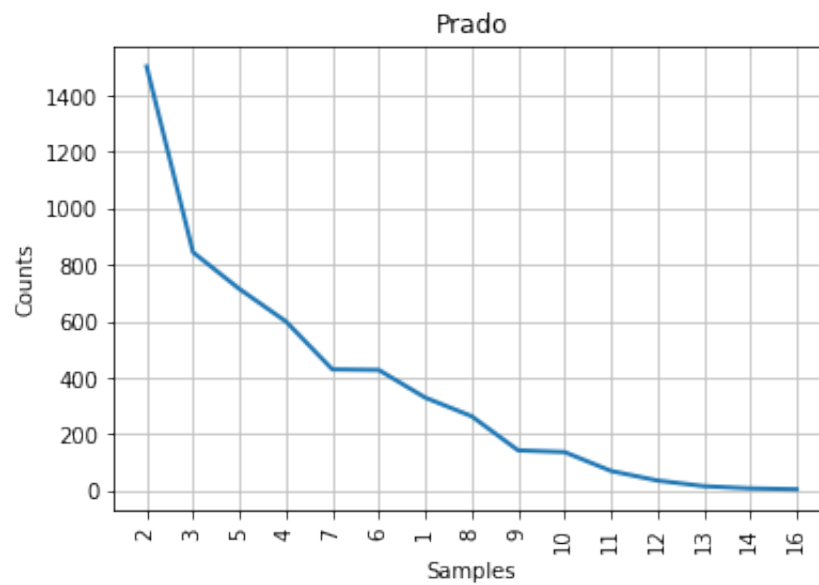


FIGURA 3.6. 1er set, 1er test, Gráfico A, Pedro Prado.

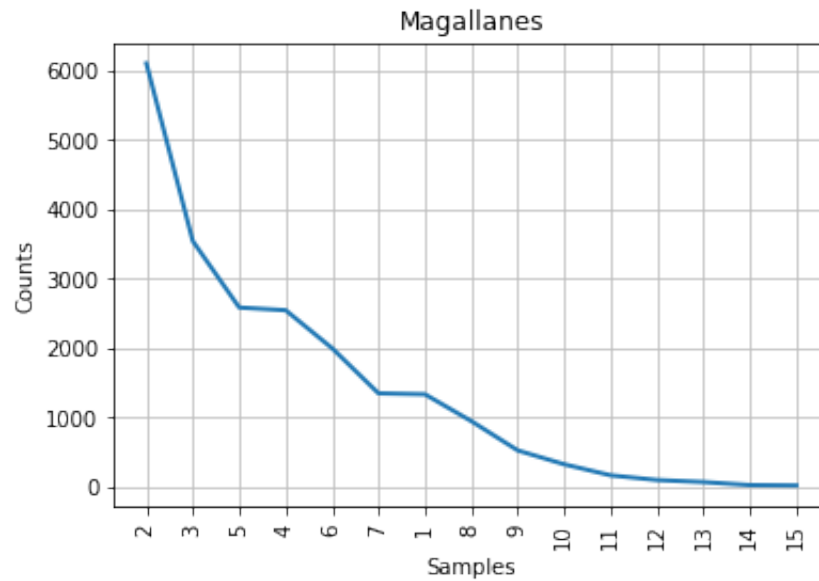


FIGURA 3.7. 1er set, 1er test,Gráfico B, Manuel Magallanes.

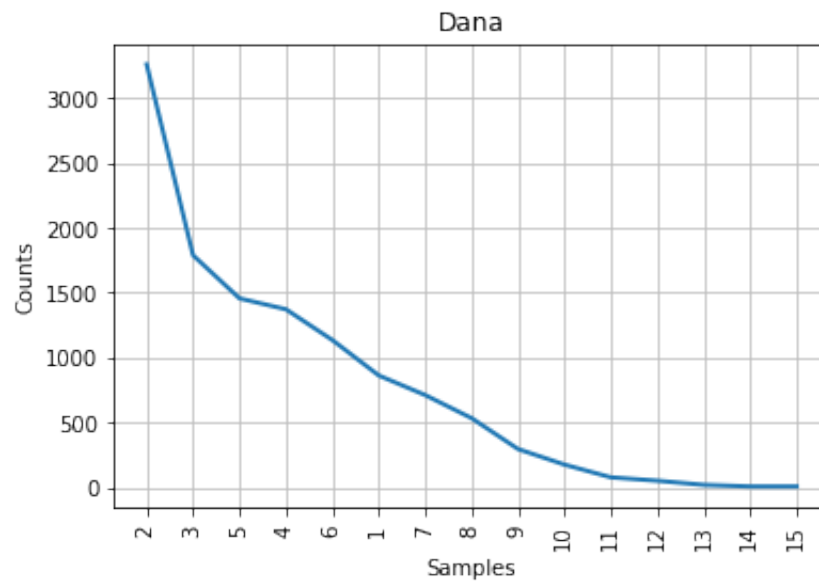


FIGURA 3.8. 1er set, 1er test, Gráfico C, Doris Dana.

En los gráficos se puede apreciar, a pesar de la diferencia de longitud debido a la diferencia de textos por corpus; que en el sector izquierdo del gráfico, todos



los documentos que representan las longitudes de palabras más frecuentes comienzan de manera similar, sin embargo cabe destacar que las curvas de Dana y Magallanes tienen una curvatura similar en el sector central izquierdo y que además el final de sus colas terminan de la misma manera. Esto es consistente con las últimas publicaciones dedicadas a la Mistral indicando su lesbianismo, su inconsciente explícita y revela a través de su epistolario lo que los medios de publicación y biografías obviaron o evitaron.

2. Segunda prueba estilométrica: Kilgariff's Chi-Squared Method.

- The Chi-squared statistic for candidate Prado is 2784.2130385443525
- The Chi-squared statistic for candidate Magallanes is 3331.2035534023858

Al estudiar la carta escrita por Mistral a Dana la prueba de Chi cuadrado nos indica que se asemeja más al set de datos enviado a Pedro Prado.

3. Tercera prueba estilométrica: John Burrows' Delta Method

- Delta score for candidate Prado is 2.563717396690494
- Delta score for candidate Magallanes is 2.347168305962745
- Delta score for candidate Dana is 2.4395963197389157

Si tomamos solo una carta para el testcase podemos apreciar una similitud respecto a la distancia de Magallanes y Dana en relación a la distancia que se ejerce con Prado, es decir el epistolario dedicado a Dana es similar al epistolario dedicado Magallanes (amatorio), que al dedicado a Prado (amistoso).

■ 2do test: Carta 39 escrita a Magallanes

1. Primera Prueba de estilometría: Utilización de curvas características de operación de Mendenhall.

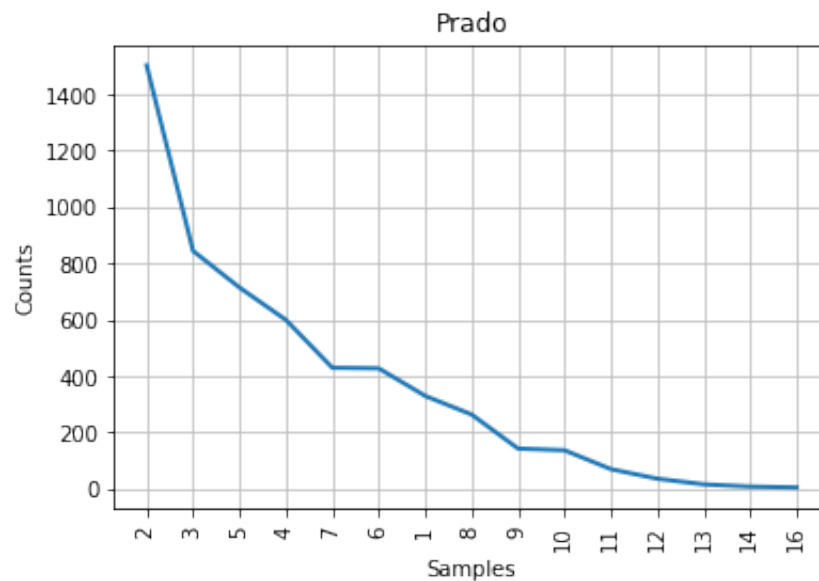


FIGURA 3.9. 1er set, 2do test, Gráfico A, Pedro Prado.

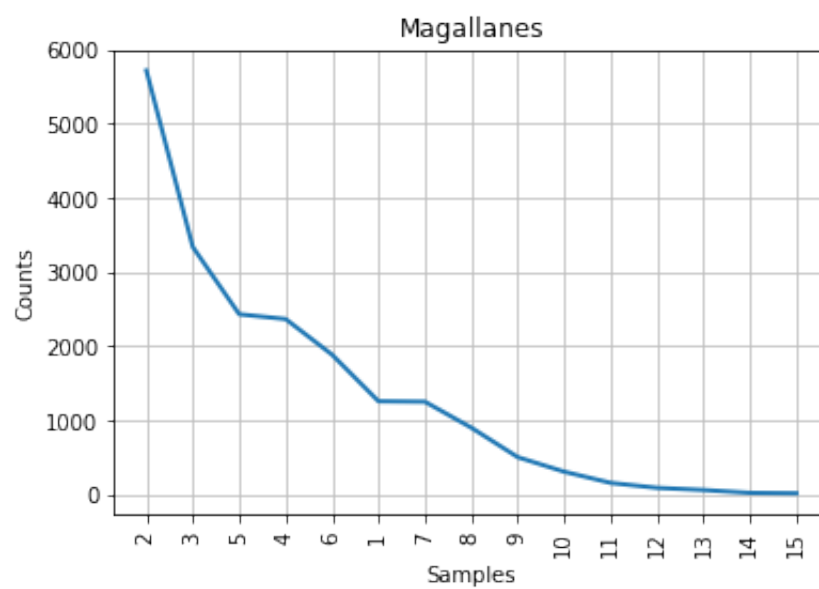


FIGURA 3.10. 1er set, 2do test, Gráfico B, Manuel Magallanes.

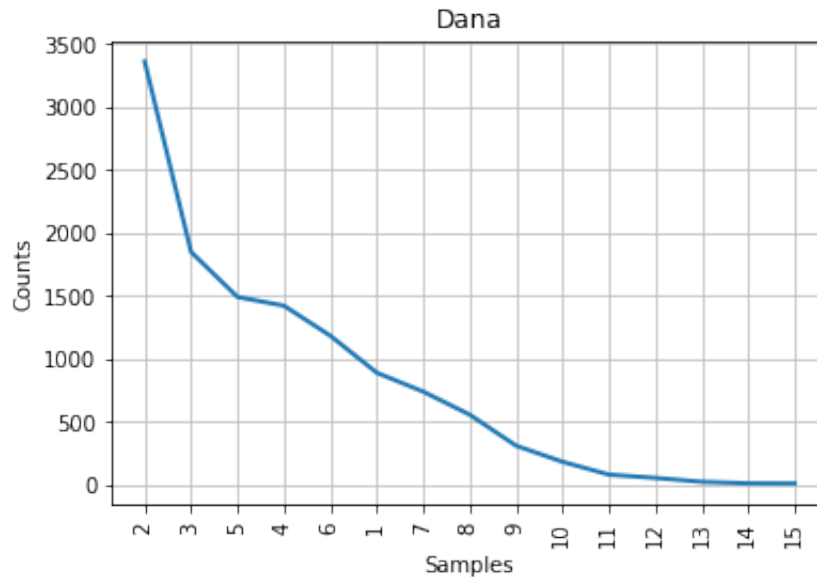


FIGURA 3.11. 1er set, 2do test, Gráfico C, Doris Dana.

## 2. Segunda prueba estilométrica: Kilgariff's Chi-Squared Method.

- The Chi-squared statistic for candidate Prado is 2782.291992745961
- The Chi-squared statistic for candidate Magallanes is 3335.960848361407

Según esta prueba el texto es similar a uno dirigido a Prado.

## 3. Tercera prueba estilométrica: John Burrows' Delta Method.

- Delta score for candidate Prado is 2.637949866030088
- Delta score for candidate Magallanes is 2.0423137853438815
- Delta score for candidate Dana is 2.3153389625198386

En este test la distribución bajó en comparación al testcase anterior, por ende se acerca más al punto de relación y certeza, lo que significa que la carta es más cercana a las escritas a Magallanes (a lo cual el software está acertando), el segundo más cercano es el corpus dedicado a Dana, relevando la similitud de corpus observada anteriormente, y en lejanía está

ubicado el corpus amistoso dedicado a Prado.

2do set de estudio:

- 1er test: Cartas 18, 25 y 38 escritas a Magallanes.
  1. Primera Prueba de estilometría: Utilización de curvas características de operación de Mendenhall.

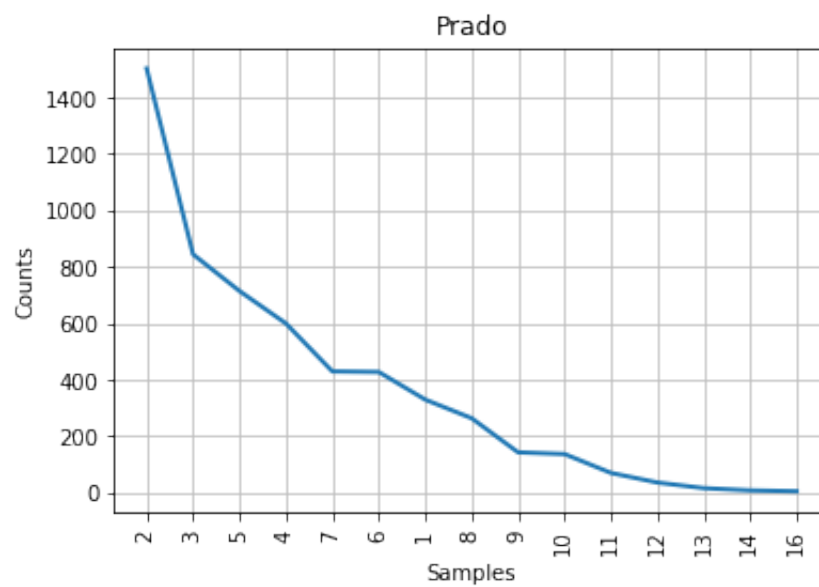


FIGURA 3.12. 2do set, 1er test, Gráfico A, Pedro Prado.

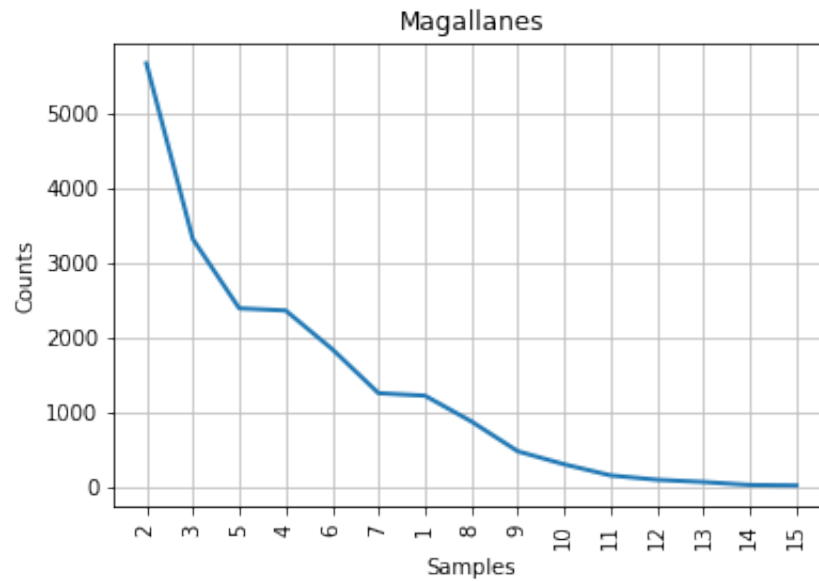


FIGURA 3.13. 2do set, 1er test, Gráfico B, Manuel Magallanes.

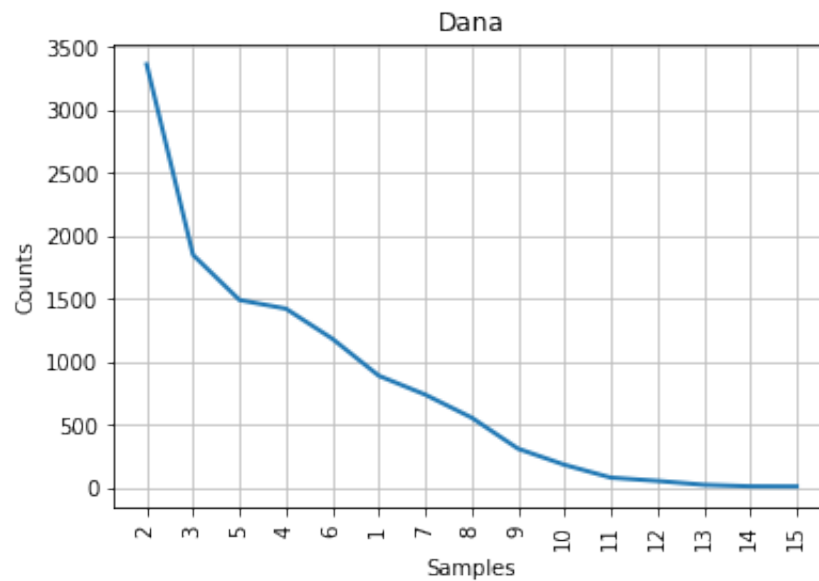


FIGURA 3.14. 2do set, 1er test, Gráfico C, Doris Dana.

A pesar de que la curva de Dana es menos angulosa que las anteriores, es visible notar la similitud entre las curvas de Dana y Magallanes, por ende podemos

observar que sus corpus son parecidos.

2. Segunda prueba estilométrica: Kilgariff's Chi-Squared Method.

- The Chi-squared statistic for candidate Prado is 2782.291992745961
- The Chi-squared statistic for candidate Magallanes is 3302.1418824575303

En el ejercicio de Chi cuadrado el resultado siempre se asimila a Prado.

3. Tercera prueba estilométrica: John Burrows' Delta Method

- Delta score for candidate Prado is 2.322161478086984
- Delta score for candidate Magallanes is 2.289278087815507
- Delta score for candidate Dana is 2.570375275055084

La tercera prueba estilométrica es certera en identificar el estilo de las cartas de Gabriela y relacionar que los documentos de prueba efectivamente son dirigidos a Magallanes.

■ 2do test: Cartas 57, 63 y 77 escritas a Dana.

1. Primera Prueba de estilometría: Utilización de curvas características de operación de Mendenhall.

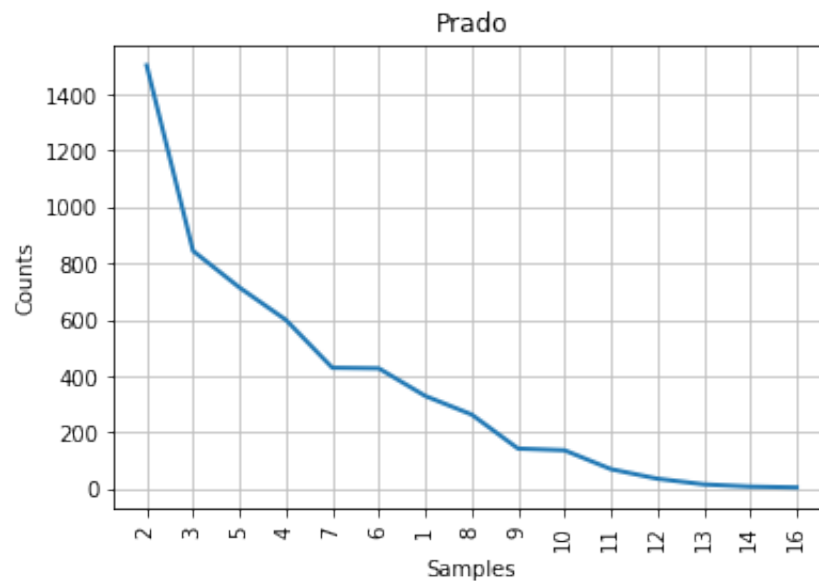


FIGURA 3.15. 2do set, 2do test, Gráfico A, Pedro Prado.

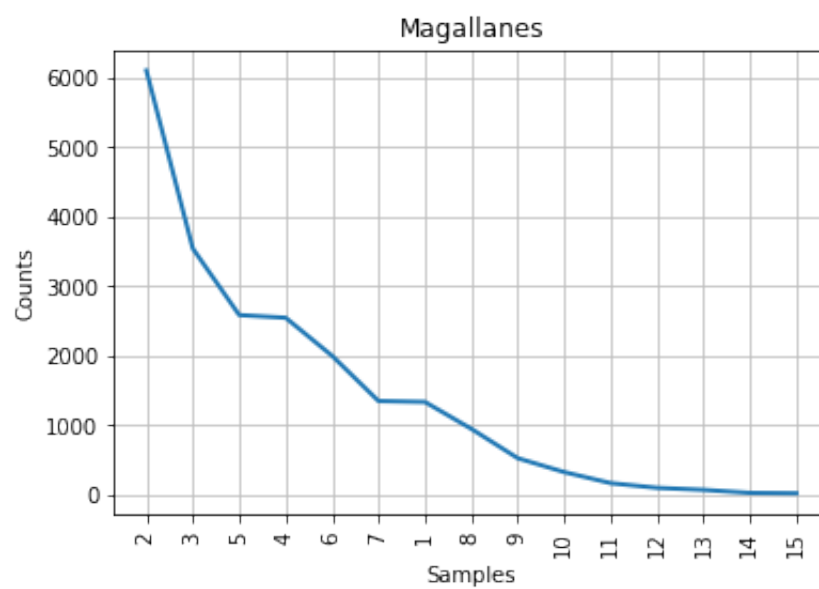


FIGURA 3.16. 2do set, 2do test, Gráfico B, Manuel Magallanes.

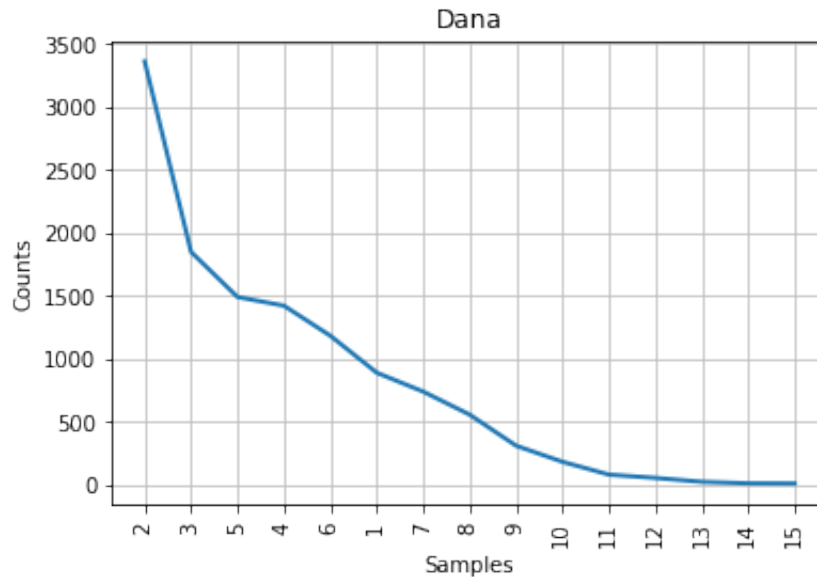


FIGURA 3.17. 2do set, 2do test, Gráfico C, Doris Dana.

En esta primera prueba, tomando como testcase un conjunto de cartas, el software muestra que las curvas de Dana y Magallanes son parecidas, más aun si nos fijamos en la cola (sector derecho inferior), que tiene un largo e inclinación parecidos.

## 2. Segunda prueba estilométrica: Kilgariff's Chi-Squared Method.

- The Chi-squared statistic for candidate Prado is 2719.376299830917
- The Chi-squared statistic for candidate Magallanes is 3235.7766784452874

El test siempre arroja un parecido al corpus destinado a Prado.

## 3. Tercera prueba estilométrica: John Burrows' Delta Method.

- Delta score for candidate Prado is 1.8877080390663745
- Delta score for candidate Magallanes is 1.4821489679139612



- Delta score for candidate Dana is 1.1910289114007477

El resultado de la distribución se acota considerablemente, lo que indica que aumentar la muestra de la prueba es un aporte al test, y verifica que el conjunto de prueba efectivamente es cercano a Dana y en segundo lugar puede asemejarse al corpus de Magallanes.

#### 4. ANÁLISIS DE RESULTADOS

De acuerdo a la muestra y a la técnica utilizada se puede demostrar lo siguiente:

Según lo calculado podemos ver que el costo computacional del Chi Square es lineal y el del delta score es cuadrático, lo que significa que el computador se demora 1 segundo en analizar las 10.000 primeras palabras, en cambio en el tercer ejercicio el computador se demoraría 10.000 segundos, ya que demoraría 10.0000 veces más, esto también refleja que para este ejemplo la tercera prueba estilométrica, la avanzada, extrajo mejores resultados.

El segundo set de estudio extrajo deltas más bajos lo que significa que se obtuvieron resultados más acertados que en el 1 set, lo que significa que fue mejor utilizar más cartas en la muestra que solo una carta.

El ejercicio propuesto por Laramée no es replicable en su totalidad en este ejercicio, ya que la segunda prueba estilométrica: Kilgariff's Chi-Squared no es aplicable al estudio de los epistolarios en verificar la semejanza de estilo estudiando los 3 corpus, solo estudia el de Magallanes y Prado, ya que está hecho y relacionado con el estudio principal de Laramée el cual se formuló para revisar los textos disputados entre 2 autores y buscando un autor oficial. Sin embargo Laramée indica que el método del Chi cuadrado es un “método burdo” ya que las formas en que los autores usan palabras más inusuales pasan inadvertidas debido a que las palabras que aparecen con mucha frecuencia tienden a tener una cantidad desproporcionada de peso en el cálculo final.

Respecto a los resultados obtenidos, se puede observar lo siguiente:

El primer set de estudio, al realizar el testeó con una carta como muestra, arroja como resultado un delta más alto, y relaciona los corpus a través de las similitudes de los corpus, o sea al estilo implicado en la autoría. En la figura (4.1), se aprecia esta conclusión de manera visual en donde a través de solo una muestra el software no resuelve la incógnita de la propiedad del subcorpus al que pertenece la muestra (carta), pero si es capaz de resolver o asimilar los subcorpus semejantes al contenido de la muestra. En este caso en particular

relaciona los corpus de Dana y Magallanes al tomar una muestra de uno y otro, pero nunca lo relaciona al de Prado.

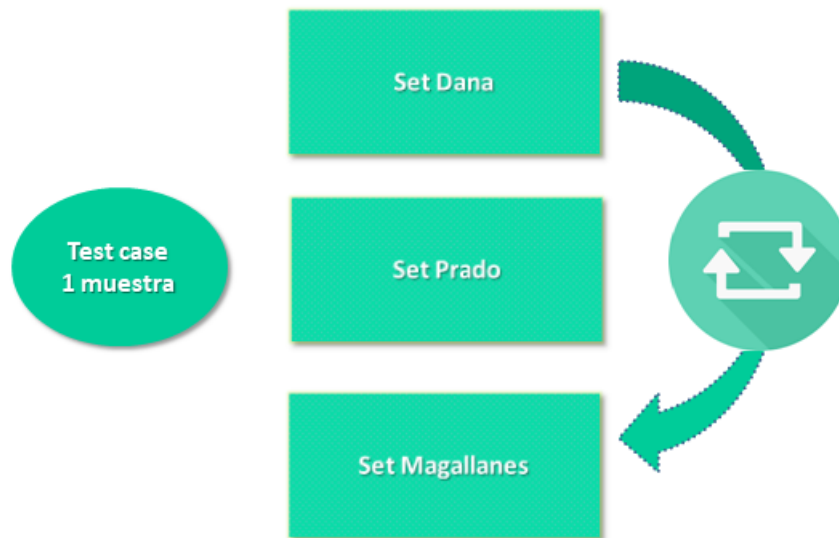


FIGURA 4.1. Utilizando 1 muestra en el test case.

El Método Delta está diseñado para comparar un texto anónimo (o un conjunto de textos) con las firmas de muchos autores al mismo tiempo. Laramée explica que delta mide cómo el texto anónimo se diferencia de los conjuntos de textos escritos por un número  $x$  de autores.

Los resultados arrojados en la tercera prueba estilométrica, del segundo set de estudio, son certeros en identificar el corpus desde donde se sacó la muestra  $x$  y deja como segunda opción el corpus más parecido al corpus identificado, explícitamente al testear una carta dirigida a Magallanes Moure, el programa identifica que la carta fue extraída de aquel corpus y da como corpus más cercano al de Magallanes Moure el de Dana y al testear al revés, el programa identifica que la carta pertenece al corpus destinado a Dana y como segundo el Delta Score de Magallanes Moure, por lo que podemos identificar que según el programa estos dos corpus se asimilan más, y ninguno de estos se asimila al corpus dedicado a Prado, como resumen de lo anteriormente expuesto se puede visualizar en la figura (4.2), que utilizando 3 muestras en un test case el programa busca y encuentra acertadamente.

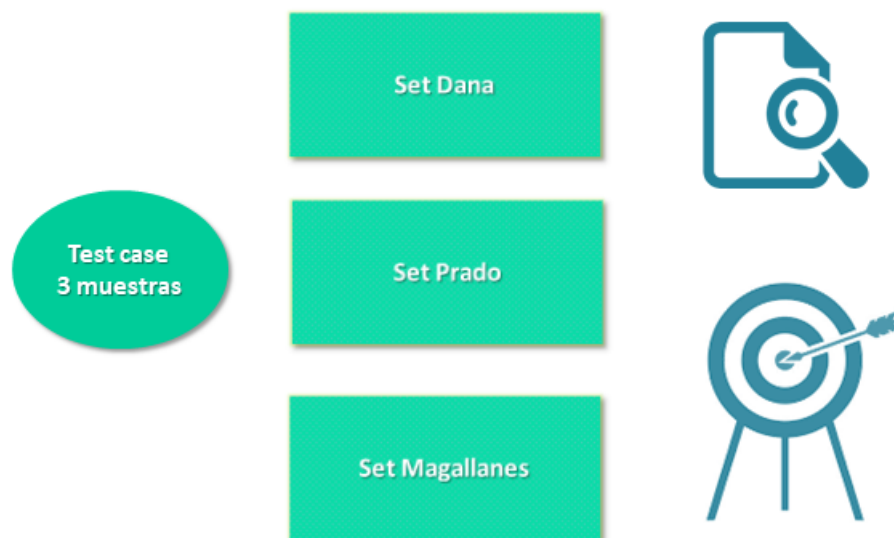


FIGURA 4.2. Utilizando 3 muestras en el test case.

El segundo test, en el que se utilizaron más cartas para aportar más datos al ejercicio, arroja un delta más pequeño y acierta al epistolario desde donde se sacaron las muestras, para este test, es mejor utilizar el Método Delta de John Burrows, ya que sus resultados aciertan al autor real. Laramée indica que “para apreciar los resultados obtenidos; cuanto menor sea la puntuación del Delta, más similar será la firma estilométrica del texto a la del candidato” sin embargo con la realización de este ejercicio también podemos indicar que a mayor cantidad de documentos en el test mejor será el resultado obtenido.

#### 4.1. Word Cloud

Para apoyar la Tercera prueba estilométrica de John Burrows’ Delta Method, del set 1, en donde se utiliza 1 muestra y el resultado tiende a asimilar los corpus dedicados a Dana y Magallanes Moure se confeccionó un wordcloud de cada corpus, basándonos en los ejercicios propuestos en la página (Mueller, 2018).

Se generó un documento .txt con el contenido de cada subcorpus para ser leído en Python y comparar de manera visual, las diferencias o similitudes de los términos utilizados en cada uno de ellos.

- Corpus Pedro Prado: Palabras destinadas a Pedro Prado.

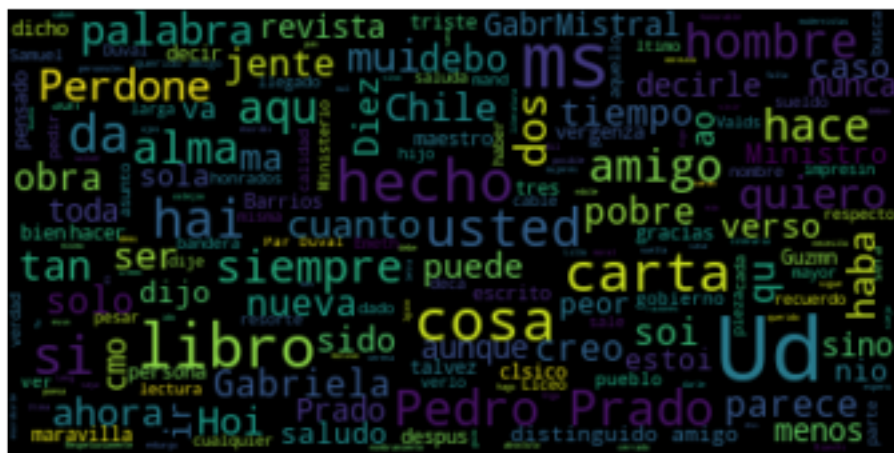


FIGURA 4.3. Nube de palabras, Pedro Prado

En la figura (4.3) podemos destacar la palabra amigo, la cual concuerda con la biografía de Mistral en donde se relaciona a estos dos personajes por la amistad que sostuvieron y que comenzó por la admiración que tenía la poetisa hacia el trabajo de Prado. Destaca la frase "distinguido amigo"

En este epistolario también se repiten los términos revista y libro, por lo cual observamos que las conversaciones sostenidas son de un tinte intelectual en donde destaca el uso de la palabra Diez que corresponde a las conversaciones en torno al Grupo de los Diez.

- Corpus Manuel Magallanes: Palabras destinadas a Manuel Magallanes Moure.





donde no se observa aquel término, pero si la palabra amigo, término no observado en los otros corpus.

También se repite el término vida, el cual no aparece en el corpus de Prado, en donde se reflejan términos relacionados al quehacer literario.

Podemos indicar que la nube de palabras apoya lo arrojado en los resultados del ejercicio del primer set, en donde relaciona y asimila los corpus de Magallanes y Dana, ya que además de la tendencia de los conectores estudiados, también nos podemos apoyar por el uso de ciertas palabras claves al comparar los términos utilizados en los corpus.

**Como conclusión de los análisis realizados podemos indicar lo siguiente respecto a los resultados:**

Respecto a la hipótesis que el ejercicio propuesto por Laramée es replicable en epistolarios de Gabriela Mistral. Tomando el trabajo realizado, a través de los estudios estilométricos se puede indicar que los ejercicios 1 y 3 propuestos por Laramée fueron replicables en nuestro trabajo, revelando aspectos importantes en los epistolarios de Gabriela Mistral y su contenido, a diferencia del ejercicio 2 que respondía a buscar una autoría teniendo en cuenta que el corpus de la lección confeccionada por Laramée contenía documentos escritos por 3 autores a diferencia del presente trabajo que pertenece a un solo autor y que por ende no hay una disputa de autoría.

La segunda hipótesis planteada en este trabajo se corroboró gracias a la aplicación de este análisis estilométrico en donde se pudo detectar rasgos estilísticos asociados a dos tópicos literarios concretos: uno referido hacia el amor pasional (o "Eros"), en donde situamos los epistolarios dirigidos a Manuel Magallanes y Doris Dana y el otro hacia la amistad (el equivalente al concepto griego de "Philos": tener una veneración o aprecio a algo o alguien), en donde claramente se sitúan las cartas dirigidas a Pedro Prado el "distinguido amigo" de Gabriela Mistral.



## 5. CONCLUSIÓN

Luego del proceso de revisión del corpus, compuesto por tres subcorpus (epistolarios) de Gabriela Mistral destinados a Pedro Prado, Manuel Magallanes y Doris Dana, utilizando Python y aplicando metodologías de Ciencias de Datos se puede concluir que el método Delta es el apropiado para analizar la documentación mencionada, este método sirvió para comprobar lo propuesto por Larameé en su artículo *Introduction to stylometry with Python*, con este ejercicio se demuestra que su propuesta efectivamente identifica autoría, y en este trabajo en particular al realizar los ejercicios variando la cantidad de cartas en la muestra, el método es capaz de identificar destinatarios y similitud de estilo entre los subcorpus.

A través de lo escrito por Gabriela Mistral en los epistolarios, se puede apreciar la evolución del escritor como persona y como autor, y constatar las diferencias de estilo en la literatura epistolar de Gabriela Mistral, el cual va cambiando acorde al destinatario. No es la misma forma de expresión cuando escribe a Magallanes Moure o Doris Dana que cuando escribe a Pedro Prado.

Enrique Valdés (Valdés, 2007) al referirse al estilo utilizado por Gabriela Mistral en sus trabajos indica lo siguiente: “algunos elementos propios de la estética y visión del mundo mistraliano, como son el énfasis, mediante los usos diversos del lenguaje, en el empleo del habla típicamente chilena y popular, como lo demuestra su preferencia por el dicho o refrán popular, el uso de hispanoamericanismos, regionalismos y coloquialismos en su discurso”. En los epistolarios analizados se puede ver reflejada la misma estética indicada por Valdés, y además se visualiza que sus escritos son expresionistas ya que ella expresa reacciones emocionales tomando un objeto como referencia. Un claro ejemplo lo constituyen las cartas dirigidas a Magallanes Moure, en donde podemos observar oraciones como las siguientes: “el cielo está amoratado y rojo: es el color de la violencia”. “Este mediodía se te asemeja mucho. Es suave y ardiente”. “El recuerdo es como las tenazas calientes y hace daño y también exaspera” Esta última frase fue dirigida a Doris Dana, 30 años después de las dos frases anteriores y vemos que el estilo expresivo se mantiene.

Gracias a la aplicación del análisis estilométrico se pueden reconocer diferencias de estilo en el corpus textual seleccionado de Gabriela Mistral, como lo es el denominar a sus destinatarios por el nombre de pila o dejar muy en claro, a través del término utilizado para dirigirse a su destinatario, la naturaleza de su relación, específicamente a Manuel Magallanes lo trata directamente de Manuel u otorgando adjetivos como Manuel amado o Poeta, a Doris Dana la llama por Doris, Dorisin o Cara, y a Prado se dirige como Pedro Prado, amigo o la frase "distinguido amigo", frase ausente en los dos epistolarios restantes en donde destacan y se repiten los términos: vida, conmigo y amor. Esta observación se condice con los gráficos y resultados de similitud anteriormente vistos en el trabajo.

Es claro distinguir, basándose en el resultado de los análisis estilométricos y lo visto en el wordcloud de esta investigación, que el epistolario dirigido a Doris Dana se sitúa en el tópico amoroso y concuerda con las publicaciones realizadas desde 1980 en adelante respecto a los contenidos de las cartas de Gabriela Mistral dirigidas a Doris Dana en donde se indica una relación amorosa entre ellas.

El aporte de este trabajo es que podemos considerarlo como el primer estudio metodológico de análisis estilométrico aplicado a epistolarios de Gabriela Mistral. En donde se estableció y determinó la naturaleza de la relación que existía entre Mistral con Pedro Prado (amistad), contrastándola con la similitud de la relación epistolar mantenida con Manuel Magallanes Moure y Doris Dana, los cuales se establecen como epistolarios de carácter amoroso, aportando al debate y a la reconstrucción de la biografía e imaginario en torno a Gabriela Mistral.

Además, mediante este trabajo se aportan datos para aumentar las cifras de estudios estilométricos latinoamericanos, se concreta un aporte chileno para las Humanidades Digitales, aporte caracterizado por estar enfocado en una figura chilena y latinoamericana, y a su vez ser un modelo de trabajo en español, utilizando epistolarios, basado en la aplicación de tecnologías de estudios estilométricos.

## BIBLIOGRAFIA

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243–257.

*Biblioteca de humanidades - sistema de bibliotecas uc.* (s.f.). <http://bibliotecas.uc.cl/biblioteca-humanidades/biblioteca-de-humanidades>. ((Accessed on 01/13/2019))

Bravo, M. (2018). Mapa estilométrico de la narrativa de eduardo mendoza: aproximación a un análisis estilístico computacional de textos literarios. *Epos: Revista de filología*(33), 99–114.

Brito, E. (2014). Una poética del género: la escritura mistraliana en niña errante. *Aisthesis*(55), 29-39.

Calvo Tello, J. (2016). *Entendiendo delta desde las humanidades*. <http://revistacaracteres.net/revista/vol5n1mayo2016/entendiendo-delta/>. ((Accessed on 01/14/2019))

Carro Paz, R. (s.f.). *Muestreo de aceptación*. [http://nulan.mdp.edu.ar/1618/1/13\\_muestreo\\_aceptacion.pdf](http://nulan.mdp.edu.ar/1618/1/13_muestreo_aceptacion.pdf). ((Accessed on 01/13/2019))

Castro, R. (1965). *Pedro prado : (1886-1952)*. Santiago, Chile: Andrés Bello.

*Chi-square test | the concise corsini encyclopedia of psychology and behavioral science - credo reference.* (s.f.). [https://search-credoreference-com.pucdechile.idm.oclc.org/content/entry/wileypsych/chi\\_square\\_test/0](https://search-credoreference-com.pucdechile.idm.oclc.org/content/entry/wileypsych/chi_square_test/0). ((Accessed on 01/13/2019))

Cuneo, A. (2017). *Conversando poemas de gabriela mistral*.

*Doris atkinson concreta traspaso a chile de material de gabriela mistral.* (2007, mayo). <https://www.emol.com/noticias/magazine/2007/05/11/255589/doris-atkinson-concreta-traspaso-a-chile-de-material-de-gabriela-mistral.html>. ((Accessed on 01/13/2019))

Facultad de Filología, U. C. d. M. (2011). *Taller sobre herramientas de análisis textual: La herramienta sketch engine.* [https://eprints.ucm.es/13796/1/SketchEngine-manual\\_nov2011.pdf](https://eprints.ucm.es/13796/1/SketchEngine-manual_nov2011.pdf). ((Accessed on 01/13/2019))

Fiol-Matta, L. (2002). *A queer mother for the nation: the state and gabriela mistral.* Minneapolis: University of Minnesota Press.

*Gabriela mistral: poeta y lesbiana.* (2003). <https://www.eltiempo.com/archivo/documento/MAM-995948>. ((Accessed on 01/13/2019))

Gelbukh, S. G. C. H. L., A. (s.f.). *Corpus virtual, virtual: Un diccionario grande de contextos de palabras españolas compilado a través de internet.* <https://www.gelbukh.com/CV/Publications/2002/MLIA-2002-Corpus.pdf>. ((Accessed on 11/13/2019))

*Guía de aprendizaje de python. programación en castellano.* (2019). [https://programacion.net/articulo/guia\\_de\\_aprendizaje\\_de\\_python\\_65](https://programacion.net/articulo/guia_de_aprendizaje_de_python_65). ((Accessed on 01/13/2019))

Jin, Y. (2017). Development of word cloud generator software based on python. *Procedia Engineering*, 174, 788 - 792. Descargado de <http://www.sciencedirect.com/science/article/pii/S1877705817302230> (13th Global Congress on Manufacturing and Management Zhengzhou, China 28-30 November, 2016) doi: <https://doi.org/10.1016/j.proeng.2017.01.223>

Laramée, D. (2018). *Introduction to stylometry with python | programming historian.* <https://programminghistorian.org/en/lessons/>

introduction-to-stylometry-with-python. ((Accessed on 01/13/2019))

lexical computing. (s.f.). *Sketch engine*. <https://www.sketchengine.eu/>. ((Accessed on 01/13/2019))

Loper, . B. S., E. (2002). *Nltk: The natural language toolkit*. <https://arxiv.org/pdf/cs/0205028.pdf>. ((Accessed on 01/13/2019))

López, J. (2012). *¿qué es la estilometría y para qué sirve?* <http://humanidadesdigitales.net/blog/2012/10/27/que-es-la-estilometria-y-para-que-sirve/>. ((Accessed on 01/13/2019))

*Magíster en procesamiento y gestión de información*. (s.f.). <https://mpgi.uc.cl/>. ((Accessed on 01/13/2019))

Mistral, G. (1978). *Cartas de amor, de gabriela mistral* (1a ed. ed.). Santiago, Chile: Andrés Bello.

Mistral, G. (1993). *En batalla de sencillez de lucila a gabriela : : (cartas a pedro prado), 1915-1939*. Santiago, Chile: Dolmen.

Mistral, G. (1999). *Cartas de amor y desamor*. Barcelona: Andrés Bello.

Mistral, G. (2005). *Manuel, en los labios por mucho tiempo : epistolario entre lucila godoy alcayaga (gabriela mistral) y manuel magallanes moure*. Santiago, Chile: Eds. Universidad Católica.

Mistral, G. (2013). *Tala* (2a. ed.. ed.). Santiago, Chile: Eds. Universidad Diego Portales.

Mueller, A. (2018). *word\_cloud: A little word cloud generator in python*. [https://github.com/amueller/word\\_cloud](https://github.com/amueller/word_cloud). ((Accessed on 01/14/2019))

Mundo, B. N. (2016, marzo). *"la prefieren loca que lesbiana": la deuda de chile con gabriela mistral, la latinoamericana que ganó el premio nobel de literatura.* [https://www.bbc.com/mundo/noticias/2016/03/160307\\_cultura\\_chile\\_gabriela\\_mistral\\_lesbianismo\\_ap](https://www.bbc.com/mundo/noticias/2016/03/160307_cultura_chile_gabriela_mistral_lesbianismo_ap). ((Accessed on 01/13/2019))

Murry, J. (1956). *El estilo literario*. México: Fondo de cultura económica.

*Natural language toolkit — nltk 3.4 documentation.* (2019). <https://www.nltk.org/>. ((Accessed on 01/13/2019))

*Personajes - memoria chilena, biblioteca nacional de chile.* (s.f.). <http://www.memoriachilena.cl/602/w3-propertyvalue-137768.html>. ((Accessed on 01/13/2019))

*Python 3.7.2 documentation.* (2019). <https://docs.python.org/3/>. ((Accessed on 01/13/2019))

Real Academia Española, D. L. E. (2019). *Definición de epistolario.* <http://dle.rae.es/?id=Fy7Qlfy>. ((Accessed on 01/13/2019))

Rodríguez, J. (2013, septiembre). *Luis vargas saavedra: "mistral fue una humanista incesante".* <http://impresa.elmercurio.com/Pages/NewsDetail.aspx?dt=08-09-2013%20:00:00&NewsID=163668&dtB=16-05-2018%20:00:00&BodyID=6&PaginaId=2>. ((Accessed on 01/13/2019))

Rohter, L. (2013). *Mother of the nation, poet and lesbian?; gabriela mistral of chile re-examined.* <https://www.nytimes.com/2003/06/04/books/mother-of-the-nation-poet-and-lesbian-gabriela-mistral-of-chile-re-examined.html>. ((Accessed on 01/13/2019))

Rossum, G. v. (2017). *El tutorial de python*. <http://docs.python.org.ar/tutorial/pdfs/TutorialPython3.pdf>. ((Accessed on 01/13/2019))

UniCo. (s.f.). *La estilometría*. <http://www.uniondecorrectores.org/la-estilometria/>. ((Accessed on 01/13/2019))

Valdés, E. (2007). *La prosa de gabriela mistral : época y estilo*. Concepción: Eds. LAR.