Análisis estilométrico para la atribución de autoría de un libelo anónimo de 1968: ¡El Móndrigo! Bitácora del Consejo Nacional de Huelga.

Dr. Gerardo Eugenio Sierra Martínez, Tonatiuh Hernández García

Grupo de Ingeniería Lingüística, Instituto de Ingeniería, UNAM.

**Resumen**: ¡El Móndrigo! es un libelo anónimo publicado en 1968 para desprestigiar al movimiento estudiantil. Para realizar la atribución de autoría, se realizó análisis estilométrico híbrido comparado entre un corpus segmentado de posibles autores y la estilometría del libelo. Se propone un método de segmentación por párrafos útil para textos complejos con más de un autor y un sistema híbrido basado en signos de puntuación y longitudes de párrafo y palabra para identificar el estilo de cada escritor. Se resolvió la atribución de autoría del libelo presentando pruebas tangibles basadas en estadística.

**Palabras clave**: lingüística forense, estilometría, movimiento del 68, atribución de autoría, estilometría comparada, 1968, similitud textual, ingeniería lingüística.

Poco después de la masacre de octubre de 1968 en la plaza de las Tres Culturas, un libro publicado por la editorial ficticia Alba Roja se repartía en los mítines del PRI, llegaba por correo a miles de domicilios particulares y se regalaba en puestos de periódicos.

El libro, de 182 páginas se titula: ¡El Móndrigo! Bitácora del Consejo Nacional de Huelga y en su primera página relata que el libro era el diario íntimo de un joven encontrado muerto al pie del edificio Chihuahua. Hallado entre sus pertenencias, daba cuenta de sucesos de máxima relevancia, puesto que había pertenecido a un líder del CNH. Nadie reconoció el cuerpo de este personaje, sólo se supo su apodo: “el Móndrigo” por lo que fue sepultado en calidad de desconocido, pero el legajo de papeles encontrado entre sus ropas fue publicado por Alba Roja debido a la importancia de su contenido. Esta anécdota es falsa.

¡El Móndrigo! no es un diario, ni si quiera una novela. Es un libelo que dedica sus páginas a difamar y desinformar sobre el movimiento del 68. Mezcla información real sobre asambleas, marchas y conflictos con mentiras, burlas y difamaciones. Los libelos son “mecanismos empleados ya sea por el Estado o por algún poder fáctico para restarle fuerza a una figura política, para deslegitimar un movimiento social o para promover una campaña de manera pérfida, entre otros objetivos no menos nefastos” (Monsiváis, 2018)

Publicado de manera anónima, no había ninguna forma de saber quién había escrito el infame libelo, la editorial Alba Roja no existe. El Grupo de Ingeniería Lingüística decidió iniciar una investigación coordinada por el Dr. Gerardo Sierra para descubrir quién fue el autor de este vergonzoso mecanismo de estado que buscaba restar credibilidad a un movimiento estudiantil meses después de la masacre.

Las siguientes líneas dan cuenta de la investigación y proceso que se llevó a cabo para determinar la autoría, combinando una investigación histórica del documento con técnicas de atribución de autoría, un área de investigación en la cual el Grupo de Ingeniería Lingüística es pionero y referente en México.

La estilometría es un campo de la lingüística forense y lingüística computacional. Se basa en el principio de que existen características textuales inconscientes y medibles que caracterizan el estilo particular de un autor. (Stamatatos, 2019)

El 5 de octubre de 2003, en una entrevista, Sergio Romero Ramírez alias El fish, porro, activista, golpeador progubernamental, admirador de Díaz Ordaz, fundador de Comando de Acción Revolucionaria Armada, vinculado al MURO y al Yunque, aporta datos sobre la elaboración de libros que “pretendieron torcer la verdad de la represión gubernamental”. (Redacción, 2003)

“De la Secretaría de Gobernación salieron estos tres libros. Todos son de la misma editorial y son de la misma pluma, un hombre de paja, porque los autores no existen

-¿Quién escribió los libros entonces?

-La Secretaría de Gobernación, los repartía la Dirección Federal de Seguridad (DFS) y los responsables eran Fernando Gutiérrez Barrios, Miguel Nazar Haro y Luis de la Barreda

-¿Pero específicamente quién se encargó de redactarlos?

-Jorge Joseph, que fue alcalde de Acapulco, por instrucciones de Gutiérrez Barrios

Joseph fue alcalde de Acapulco a principios de los sesenta Muerto hace dos años, Joseph, dice El Fish, escribió esos libros con base en los informes de la DFS, con la que colaboraban los militares Arturo Acosta Chaparro y Humberto Quirós Hermosillo” (Redacción, 2003)

¡El Móndrigo! fue escrito por órdenes del capitán Fernando Gutiérrez Barrios, quién era director de la Dirección General de Seguridad (y agente de la CIA). El encargado de escribir el libelo fue Jorge Jóseph Piedra, un periodista veracruzano que formaba parte de un grupo de agentes de inteligencia que cubrían marchas, asambleas y actividades estudiantiles en Ciudad Universitaria y otras sedes. Piedra destacó con sus informes, lo cual lo llevó a encargarse de la escritura del Móndrigo. Su destacada participación en este proyecto lo convertiría en analista político de Gobernación, en dónde fue escalando puestos hasta llegar a ser alcalde de Acapulco. (Munguía, 2016)

Jorge Joseph utilizó los archivos policiales de los agentes a su cargo para dar verosimilitud al texto. (Tasso, 2018) Mezclándolos con su prosa, crea la ilusión de un diario meticuloso. Estos informes se pueden identificar en el texto por la mención de trayectorias de marchas, ubicación de asambleas, uso abundante de nombres propios de oradores, maestros y líderes del movimiento, cifras de heridos, detenidos, muertos y desaparecidos, detalles sobre la duración y temas de las asambleas, y cifras de asistentes a mítines y marchas.

Estos informes policiales se pueden consultar en el Archivo General de la Nación, en la galería 1, con número de archivo DFS 11-4-68 L24 F131; 68/08/01; F133, F234, F236 y F241

Para terminar de complicar las cosas, aparece una tercera pluma: la de Emilio Uranga. Gerardo Medina Valdés, ex director del periódico La nación, publicó en su libro Operación 10 de junio que Emilio Uranga era el autor del Móndrigo. A esta opinión se suman Juan Miguel de la Mora, y Gonzalo Martré. (Martré, 1998)

En una carta que Octavio Paz escribió a Carlos Fuentes el 3 de agosto de 1969 se indica a Emilio Uranga. “... se promueven y se publican ascos como El Móndrigo, un folleto obra de la cucaracha llamada [Emilio] Uranga.” (Sheridan, 2018)

A pesar de que ninguno de los autores mencionados presenta pruebas tangibles que permitan asociar a Uranga con la escritura del Móndrigo, lo cierto es que, en aquel entonces, el filósofo fungía como ideólogo para la Secretaría de Gobernación. Es probable que, debido a su rango, (superior al de Joseph Piedra e incluso al de Gutiérrez Barrios,) haya tenido que encargarse de la revisión final del libelo, contando con la libertad de agregar o modificar el libelo y ello provocó los rumores de sus rivales.

Sostengo la siguiente hipótesis: ¡El Móndrigo! fue escrito por Jorge Joseph, quién utilizó informes policiales para dar verosimilitud al texto. Emilio Uranga, revisó el libelo agregando y modificando lo que juzgó necesario. La participación de Uranga en el libelo sería breve, párrafos agregados a su elección e incluso páginas enteras, pero en una proporción mucho menor a la de Joseph.

Decidí usar estilometría comparada para resolver este problema de atribución autoral. La estilometría es “un método por el que se puede optar para detectar similitud textual. No se limita a encontrar similitud exacta entre fragmentos textuales, sino que representa una opción en la búsqueda de similitud y copia cuando se ha utilizado un procedimiento avanzado para el plagio, como la paráfrasis o plagio de ideas...” (González, 2011)

Debido a la complejidad del problema, opté por este método por ser el más preciso y porque me permitirá identificar a los autores involucrados. Para establecer mis parámetros identifiqué aquellos que los autores “plasman de manera inconsciente” (González, 2011), y que sean “ajenos al género textual y a la temática de los textos” (González, 2011). Estos parámetros se conocen como marcadores estilísticos, que incluyen “marcadores el formato del texto, abreviaturas, puntuación, sintaxis, discurso y errores y corrección” (González, 2011).

Para comprobar nuestra hipótesis, abordamos al libelo como un problema de atribución de autoría, dónde midiendo distintas características textuales entre textos podemos diferenciar entre dos autores (Stamatatos, 2019).

En cada atribución de autoría se tiene un conjunto de autores candidatos, un conjunto de textos de autoría probada de cada autor candidato denominado corpus de entrenamiento y un conjunto de textos de autoría desconocida (por atribuir) llamado corpus de prueba. (Stamatatos, 2019)

Para nuestro corpus de entrenamiento formamos un conjunto de cinco autores: Emilio Uranga, Jorge Joseph Piedra, Roberto Blanco Moheno, Gregorio Ortega Molina y Gregorio Ortega Hernández. Añadiendo tres autores a la lista, aumentamos la exactitud de los resultados.

El Grupo de Ingeniería Lingüística reunió los textos de los autores además del texto completo de ¡El Móndrigo!, los cuales fueron transcritos por becarios para poder realizar los análisis pertinentes.

Los autores son contemporáneos entre sí, sus registros lingüísticos son similares y los textos son próximos temporalmente. Los textos del corpus de entrenamiento son los siguientes:

**Emilio Uranga**:

“Universidad Popular”. La Prensa, 20 de julio de 1968, pp. 3 y 35

“La Represión Extremada”. La Prensa, ¿? de ¿? de 1968, pp. 3 y 38

“La Ambigüedad Universitaria”. La Prensa, ¿? de ¿? de 1968, pp. 3 y 38

“Ornato y Orden”. La Prensa, ¿? 1968, pp. 3 y 47

“Viaje del Canciller”. La Prensa, ¿? 1968, pp. 3 y 32

Prólogo de Astucias literarias, México, Federación Editorial Mexicana, 1971, p. 9

De Astucias literarias, México, Federación Editorial Mexicana, 1971, p. 31-35 (8-I-70)

De Astucias literarias, México, Federación Editorial Mexicana, 1971, p. 228-233 (11-II-70)

Lectura de galeras de De Astucias literarias, México, Federación Editorial Mexicana, 1971, p. 299-303

**Jorge Joseph**:

“Picaluga Vende a los Rojo Gomistas por una Curul Senatorial”, de El ministro del odio, México, Edición del autor, 1961, pp. 35-40

“Pitoloco en Palacio”, de El ministro del odio, México, Edición del autor, 1961, pp. 119-124

“Aviso”, de El ministro del odio, México, Edición del autor, 1961, p. 125

“Pruebas de la existencia de la Atlántida”, de México: cuna de la civilización universal, México, Ramírez Editores, 1965, pp. 87-96

“Esoterismo en Anahuac”, de México: cuna de la civilización universal, México, Ramírez Editores, 1965, pp. 263-277

**Gregorio Ortega Hernández**:

“México resistió a la C.I.A. y a Fidel Castro”, Revista de América, 2 de noviembre de 1968.

“Un chamaco de Tepito da una lección de civismo”, Revista de América, 27 de abril de 1968.

“Un injusto artículo de Blanco Moheno”, Revista de América, # de julio de 1968. #1176

Gregorio Ortega Molina:

“¿Hacia la extrema derecha?”, Revista de América, México, 1968, 21 diciembre 1968. #1199

“La guerra de los zombies”, Revista de América, México, 1968, ¿ octubre 1968. #1191

“La sociedad del átomo”, Revista de América, México, 1968, ¿ julio 1968. #1179

**Roberto Blanco Moheno**:

“Tlatelolco. Historia de una infamia”, México, 1969, Editorial Diana, pp. 193-206

“La noticia detrás de la noticia”, México, 1966, Edición de autor, pp. 103-105

“La noticia detrás de la noticia”, México, 1966, Edición de autor, pp. 180-190

“La noticia detrás de la noticia”, México, 1966, Edición de autor, pp. 208-211

Los corpus fueron segmentados en párrafos numerados progresivamente para su identificación posterior. Buscando características que “presenten una importante variación iter-autor y poca variación intra-autor” (González, 2011) y que queden “fuera del control consiente del autor” (González, 2011) decidí usar las siguientes: longitud de párrafo (número de líneas por párrafo), número de palabras por párrafo, número de puntos por párrafo, número de comas por párrafo, número de punto y coma y dos puntos.

Estas características, por lo común suelen ser elecciones inconscientes del autor que responden a su manera de pensamiento o forma de expresión (por ejemplo, alguien con un nivel educativo mayor usa oraciones más, párrafos más largos y nexos para coordinar las oraciones, y es poco común que alguien cuide el número de comas que usa en un párrafo) por lo que he estimado que la información obtenida será valiosa.

Los análisis estilométricos de cada autor se pueden consultar en la siguiente carpeta de Google Drive:

<https://drive.google.com/drive/u/0/folders/15jpfmmAR0ISa6CsmVhZrCcNBcjPDINik>

Para identificar las características propias de cada autor, realicé el análisis estilométrico de todo el Móndrigo. Para realizarlo se limpió el texto (el libelo contaba con marcas dentro del texto, por ejemplo) numeré por párrafos para obtener las mismas características que obtuve de los corpus de entrenamiento (longitud de párrafo, número de palabras por párrafo, número de puntos por párrafo, número de comas por párrafo, número de punto y coma y dos puntos) para hacer una comparación.

Segmentar el texto en párrafos para su análisis fue una decisión que tomé por dos razones: en primer lugar, considero que el párrafo es una unidad completa de pensamiento cuya elección es inconsciente, así como la distribución de los signos de puntuación en su interior. Por tanto, dan buenos indicios del autor, pues estas medidas no varían mucho dentro de un corpus del mismo autor (consciente de que esto puede variar con el tiempo, escogí textos cercanos temporalmente a la elaboración del libelo) pero sí cambian de un autor a otro drásticamente.

Por ejemplo, la longitud de párrafo favorita de Uranga es de 9 líneas, mientras que la piedra es 5 líneas. Mientras que Piedra no usa menos de 67 palabras por párrafo, Uranga no baja de 116 palabras por párrafo. Un párrafo de Piedra nunca pasará de 166 palabras, mientras que los de Uranga llegan a 158.

Las diferencias son grandes y nos permiten marcar un estilo propio de cada autor mediante la comparación de las características. El número más común de comas que usa Piedra es 2, mientras que Uranga usa 8 comas por párrafo con una frecuencia sorprendente.

Cada una de estas características particulares fue rastreada en el Móndrigo para encontrar a cada uno de los autores que han intervenido en él.

La segunda razón para segmentar en párrafos ha sido que hay demasiados autores para hacer generalizaciones. Eso se debe a que la escritura de Uranga está mezclada con la Piedra, a veces el filósofo pone un párrafo antes de tres del periodista, luego hay un discurso y otro párrafo de Uranga seguido de un informe de inteligencia y que luego de tres hojas seguidas de Piedra encontramos un poema de Miguel Hernández. Analizarlo como un todo fue un error que no cometí.

Los informes policiales. El libelo está plagado de ellos. Mientras nos dan acceso al archivo (que se encuentra en el Archivo General de la Nación) los he identificado por las siguientes características que contienen: enumeraciones de nombres propios, enumeraciones de participantes, rutas de manifestaciones, descripción de asambleas, descripción de enfrentamientos, cifras de detenidos, muertos y heridos, cifras de asistentes a marchas, asambleas y manifestaciones, cifras de militares y policías, detalles de vehículos e inmobiliario urbano dañado.

El análisis estilométrico de ¡El Móndrigo! completo se puede consultar en el siguiente enlace.

<https://docs.google.com/spreadsheets/d/18CKkHQQfTcghbxqKAIGnzBgSCiyTcvkQu6OKOpfRrH8>

Todos los datos estilométricos se guardaron en formato .xlsx para trabajar con ellos. Utilicé Python para analizar los datos, principalmente la librería Pandas y Matplotlib. A continuación, describiré el código utilizado y presentaré los resultados obtenidos del análisis estadístico.

En primer lugar importamos las librerías que serán usadas durante el proceso:

%matplotlib inline

import pandas as pd

import matplotlib.pyplot as plt

Después, cargamos los archivos .xlsx con la función pd.read\_excel

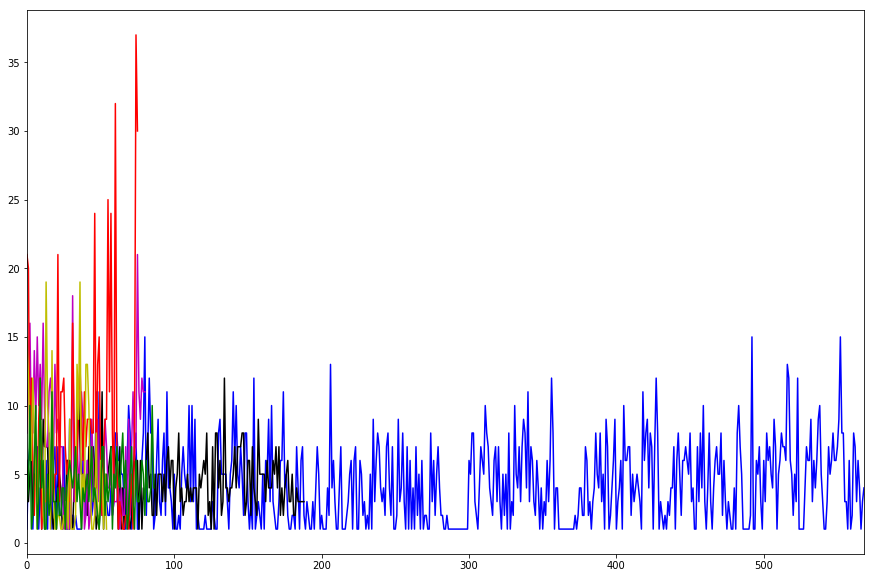
data = pd.read\_excel('UranguitaE.xlsx') #Contiene la estilometría del corpus de Emilio Uranga

data2 = pd.read\_excel('JorgitoE.xlsx') #Contiene la estilometría del corpus de Jorge Joseph Piedra

data3 = pd.read\_excel('MondriE.xlsx') #Contiene la estilometría del Mondrigo

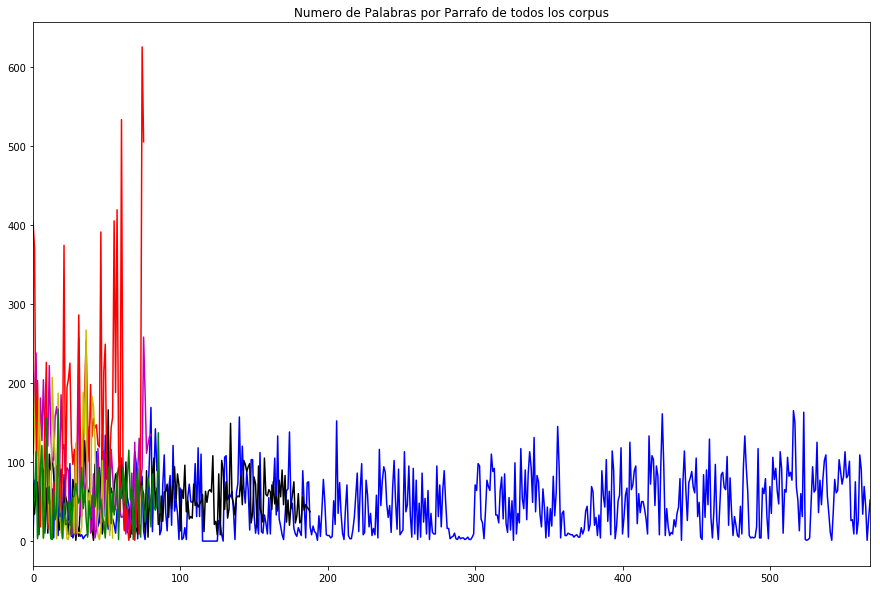
Las variables data, data2 y data3 ahora contienen los análisis estilométricos de cada corpus. Variables similares contienen el resto de los datos de los autores.

Se graficaron todas las longitudes de párrafo para ver el comportamiento de los datos:

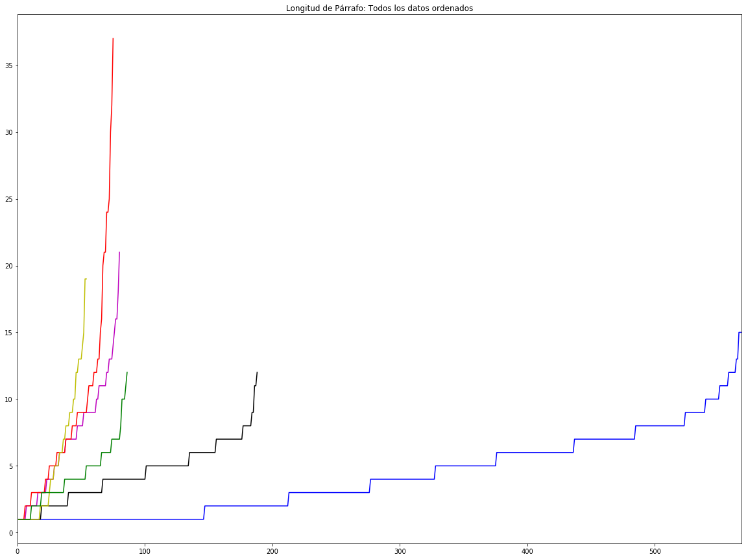


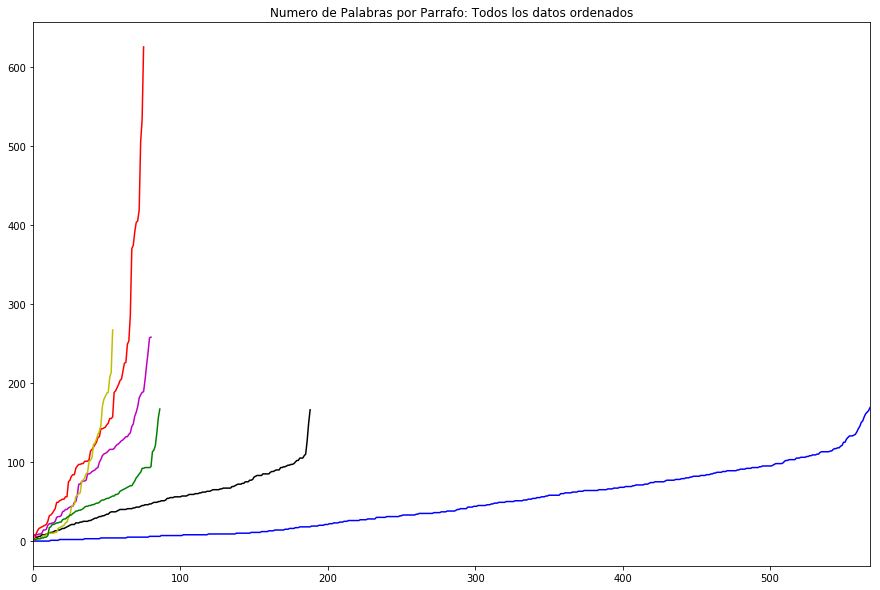
En azul vemos el Móndrigo, en negro a Jorge Joseph, en magenta a Emilio Uranga, a Moheno en rojo, Ortega en amarillo y Molina en verde.

Mostramos también la gráfica de Palabras por Párrafo:



Para visualizar mejor las similitudes y diferencias entre los corpus, se ordenaron los datos:





Se extrajeron los conteos de cada característica de los corpus de entrenamiento y del texto anónimo. Debido a la disparidad de tamaño de los corpus, se compararon las características y no la frecuencia de estas.

Se compararon las características de los corpus de autoría probada con ¡El Móndrigo! para ver el grado de similitud:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Autor | Long de párrafo | Palabras por párrafo | Puntos por párrafo | Comas por párrafo | Punto y coma x P | Dos puntos x P |
| Jorge Joseph | 78.57 | 70.54 | 72.72 | 68.42 | 55.55 | 50.00 |
| Emilio Uranga | 71.42 | 51.16 | 72.72 | 68.42 | 44.44 | 75.00 |
| Blanco Moheno | 42.85 | 50.38 | 36.36 | 26.31 | 33.33 | 75.00 |
| Gregorio Ortega | 78.57 | 35.65 | 81.81 | 68.42 | 55.55 | 50.00 |
| Ortega Molina | 78.57 | 47.28 | 63.63 | 47.36 | 33.33 | 25.00 |

Se señalan en cada columna los dos resultados más altos del análisis. Entre más se acerque el porcentaje al 100% mayor es la similitud. Jorge Joseph es quién más similitud tiene con ¡El Móndrigo!; su similitud es alta en 5 de seis características estilométricas.

Gregorio Ortega tiene 4 de 6 coincidencias, al igual que Emilio Uranga. Gregorio Ortega tiene mejores porcentajes de ajuste, pero en la categoría Palabras por párrafo tiene el porcentaje más bajo de todos los autores.

Esto es importante, pues si sus párrafos contienen menos palabras que ¡El Móndrigo! quiere decir que su estilo de escritura (párrafos cortos) es un 65% más corta que la de quién escribió el diario. Por otra parte, su similitud con respecto al uso de puntos por párrafo es la más alta de todos los autores, lo cual quiere decir que Gregorio Ortega usa párrafos más cortos, con una cantidad de puntos muy alta. Esto discrepa con el estilo de escritura de quién escribió el Móndrigo.

Emilio Uranga tiene porcentajes de similitud ligeramente más bajos (este desbalance es causado por el elevado uso de puntos de Ortega), pero son equilibrados (no hay desviaciones absolutas, como el descenso abrupto a 36% de Ortega en párrafos por palabra) en incluso supera a Ortega en dos puntos por párrafo. Desde mi punto de vista lingüístico Emilio Uranga ocuparía el segundo lugar de similitud y Gregorio Ortega el tercero.

Dadas las discrepancias entre ¡El Móndrigo! y el corpus de Gregorio Ortega, descartaría una relación.

¡El Móndrigo! fue escrito por Jorge Joseph Piedra y la evidencia indica que Emilio Uranga agregó algunos fragmentos y por ello su estilo se encuentra presente en el texto, aunque en menor medida que Jorge Joseph y por ello sus porcentajes de similitud obtenidos son bajos, pero consistentes.

Comprobamos la hipótesis, lo cuál tiene relevancia en el aspecto histórico pues la literatura sobre el libelo no ofrece pruebas sobre su autoría, cuestión que permanece sin resolución entre los especialistas, aquí, gracias a la tutoría y coordinación del Dr. Gerardo Sierra logramos determinar la autoría con pruebas matemáticas y tangibles.

Los resultados concuerdan con la investigación histórica, lo cuál es una prueba de que, trabajando en conjunto, la la lingüística y la computación tienen enormes probabilidades de hacer frente a con éxito a cualquier reto: eso es la Ingeniería Lingüística.

Agradezco al Grupo de Ingeniería Lingüística el apoyo y la capacitación brindada. Al Dr. Gerardo Sierra por las invaluables tutorías y guía: sin su capacidad de formar capital humano de primer nivel y desarrollar investigación multidisciplinaria este trabajo no hubiera sido posible. Agradezco al personal del GIL que colaboró en la transcripción de los documentos.

# Referencias

González, A. R. (2011). *Análisis estilométrico para la detección de plagio.* Ciudad de México: UNAM.

Martré, G. (1998). *El movimiento popular estudiantil de 1968 en la novela mexicana.* México: UNAM.

Monsiváis, C. (8 de 12 de 2018). *De libelos y libros*. Obtenido de https://www.proceso.com.mx: https://www.proceso.com.mx/137972/de-libelos-y-libros

Munguía, J. R. (2016). *La otra guerra secreta: los archivos prohibidos de la prensa y el poder.* México : Random House.

Redacción. (5 de octubre de 2003). *Las confesiones de El Fish*. Obtenido de https://www.proceso.com.mx: https://www.proceso.com.mx/190338/las-confesiones-de-el-fish

Sheridan, G. (04 de 12 de 2018). *Letras Libres* . Obtenido de Paz y Fuentes: cartas tlatelolcas (“el sector intelectual”): https://www.letraslibres.com/mexico/literatura/paz-y-fuentes-cartas-tlatelolcas-el-sector-intelectual

Stamatatos, E. (14 de 1 de 2019). *A Survey of Modern Authorship Attribution Methods.* Obtenido de citeseerx.ist.psu.edu: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.440.1634&rep=rep1&type=pdf

Tasso, P. (04 de 21 de 2018). *La isla sin Este*. Obtenido de El mondrigo, la novela de estado : http://laislasineste-analisis.blogspot.com/2007/11/el-mndrigo-la-novela-de-estado.html