xAtribución de autoría de *¡El Móndrigo! Bitácora del Consejo General de Huelga.*

Resultados condensados.

Dr. Gerardo Eugenio Sierra Martínez, Ing. Brian Alonso Sánchez Vázquez, Ing. Samuel Ramos Sosa, Tonatiuh Hernández García

Este trabajo tiene como propósito atribuir la autoría de un libelo anónimo publicado en 1968 por la editorial fantasma Alba Roja. El documento, de 182 páginas narra a manera de diario la supuesta historia del Móndrigo, líder ficticio del Consejo Nacional de Huelga.

El diario apócrifo fue un mecanismo empleado por el Estado Mexicano para deslegitimar el movimiento estudiantil (Monsiváis, 2018) meses después de la masacre del 2 de octubre, mediante difamaciones y mentiras.

¡El Móndrigo! fue escrito por órdenes del capitán de la Dirección Federal de Seguridad y agente de la CIA Fernando Gutiérrez Barrios. El encargado de redactar el libelo fue Jorge Joseph Piedra, un agente de la DFS que coordinaba un grupo de agentes de campo que espiaba marchas y asambleas universitarias (Redacción, 2003).

Jorge Joseph utilizó los archivos policiales de los agentes a su cargo para dar verosimilitud al texto, rescribiéndolos para crear la ilusión de un diario meticuloso (Tasso, 2018).

Emilio Uranga, filósofo e ideólogo al servicio de la Secretaría de Gobernación también intervino el libelo (Martré, 1998), (Sheridan, 2018): fue comisionado por el capitán Barrios para revisar la efectividad del libelo y agregar párrafos especialmente insidiosos.

La hipótesis de este trabajo es que ¡El Móndrigo! fue escrito por Jorge Joseph Piedra en su mayor parte, con fragmentos de longitud variable agregados por Emilio Uranga.

Para comprobarla, abordaremos 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. 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. #1193

“Un chamaco de Tepito da una lección de civismo”, Revista de América, 27 de abril de 1968. #1168

“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

El corpus de prueba está formado por:

¡El Móndrigo! Bitácora del Consejo Nacional de Huelga, México, 1968, Ed. Alba Roja, Anónimo 182 p.

A continuación, desarrollaremos distintos métodos para comprobar nuestra hipótesis. En caso de tener exito tendremos pruebas para ligar a ambos autores con el libelo exhibiendo las practicas difamatorias y de desinformación del Estado Mexicano a través de sus órganos de seguridad y gobernación. Resolveríamos también una incógnita, pues la literatura actual sobre el libelo no ha resuelto el problema de la autoría: algunas señalan a Jorge Joseph (Redacción, 2003) (Tasso, 2018), otras a Uranga sin aportar pruebas concretas (Martré, 1998) (Munguía, 2016) (Sheridan, 2018), pero ninguna ofrece nuestro enfoque de escritura colaborativa entre ambos personajes con pruebas basadas en métodos computacionales y estadísticos.

**Atribución de autoría con Support Vector Machines mediante clasificación multiclase “one vs rest” utilizando los vectores de características del Sistema Automático de Estudios Estilométricos (SAUTEE) con predicción para cada tipo de kernel.**

Se hizo una clasificación de los textos de entrenamiento mediante quince marcadores estilométricos que se convirtieron en vectores (Paass, 2003) mediante LIBSVM, implementando el paradigma OVR. Se hicieron 100 iteraciones por kernel, contemplando cuatro funciones: lineal, polinomial, rbf y sigmoidal. Se obtuvieron los mejores parámetros del clasificador, seleccionando las funciones y parámetros de kernel de mayor precisión. En caso de empate, se seleccionó aquella con función de penalización C, para dar mayor holgura al clasificador. Se hizo una tabla de predicción de autoría de los corpus de entrenamiento sobre el corpus de prueba:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Marcador Estilométrico | Kernel | Ortega Molina | Emilio Uranga | Blanco Moheno | Ortega Hernández | Jorge Joseph |
| Bigramas de caracteres | LIN | 10.61% | 3.94% | 1.82% | 14.01% | 69.62 |
| Trigramas de carácter | RBF | 8.28% | 9.43% | 4.25% | 11.72% | 66.31 |
| Unigramas de palabras funcionales | POLY | 10.05% | 8.85% | 6.90% | 13.30% | 60.91 |
| Bigramas de palabras funcionales | SIG | 16.47% | 17.45% | 12.10% | 14.24% | 39.48 |
| Bigramas de palabras funcionales con hasta 2 huecos | SIG | 9.67% | 24.94% | 14.59% | 9.19% | 42.19 |
| Trigramas de palabras funcionales | LIN | 9.02% | 29.94% | 11.41% | 10.36% | 39.26 |
| Trigramas de palabras funcionales con hasta 2 huecos | RBF | 9.16% | 35.30% | 15.41% | 17.47% | 22.66 |
| Longitud de oraciones | SIG | 14.70% | 42.58% | 8.80% | 12.37% | 21.55 |
| Longitud de palabras | RBF | 9.40% | 18.15% | 3.58% | 14.15% | 54.72 |
| Unigramas de etiquetas POS | LIN | 15.36% | 3.43% | 7.54% | 14.02% | 59.65 |
| Bigramas de etiquetas POS | POLY | 19.59% | 13.22% | 13.96% | 16.44% | 36.79 |
| Trigramas de etiquetas POS | POLY | 32.08% | 8.08% | 14.42% | 29.59% | 15.83 |
| Cat. Gram al inicio de la oración | SIG | 6.73% | 44.50% | 9.35% | 19.41% | 20.01 |
| Cat. Gram al final de la oración | LIN | 13.10% | 36.80% | 5.99% | 17.28% | 26.83 |
| Unigramas de et. POS no fino | RBF | 13.04% | 12.51% | 15.63% | 12.12% | 46.71 |

Según la información de la tabla, es probable que ¡El Móndrigo! haya sido escrito por Jorge Joseph con contribuciones de Emilio Uranga.

**Método Kilgarrif**

Propuesto por Adam Kilgarriff en (Kilgarriff, 2001) establece la medida de “distancia”: si un autor escribe dos textos y estos se comparan la distanciadebe ser muy pequeña, para concluir que se tratan del mismo autor. Se compararon las 500 palabras más comunes del corpus de cada autor con ¡El Móndrigo! siguiendo la siguiente ecuación: =

**Método Delta de John Burrows**:

John Burrows propone en (Burrows, 2002) un método que, dado un documento de prueba y una serie de documentos cuya autoría está plenamente definida, arroja una medida de distancia con todos los autores, identificando un estilo particular de cada autor y midiendo la distancia entre ese estilo y el texto en disputa, el de prueba. Esta medida se llama Δ y entre más pequeña más certeza habrá del origen de ese texto.

|  |  |  |
| --- | --- | --- |
| **Autor** | **Distancia** | **Delta Δ** |
| Emilio Uranga | 1771.0364889378193 | 1.5134148360747048 |
| Gregorio Ortega Molina | 2002.4749684727008 | 1.0422030585665234 |
| Jorge Joseph | 1643.598487830984 | 0.5190645312503361 |
| Gregorio Ortega Hernández | 1986.895626691162 | 1.2069074912312774 |
| Roberto Blanco Moheno | 3221.980751878392 | 1.2278603090115858 |

Jorge Josehp y Emilio Uranga son los más cercanos al Móndrigo de acuerdo a la distancia . La medida DELTA señala a Jorge Joseph como el autor más probable, pero descarta segundas opciones.

**Estilometría híbrida**

Se diseñó una aproximación híbrida basada en signos de puntuación por párrafo (punto, coma, punto y coma, dos puntos), longitud de párrafo y número de palabras por párrafo. Este tipo de aproximación difiere de las medidas estilométricas típicas.

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 es plausible suponer 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.

**Conclusiones:**

El análisis con máquinas de soporte vectorial indica que Jorge Joseph Piedra es quién tiene más coincidencias con ¡El Móndrigo! seguido de Emilio Uranga y con una similitud muy baja para Gregorio Ortega.

El análisis con el método de Kilgarriff señala a Jorge Joseph y a Emilio Uranga como los autores más probables de ¡El Móndrigo!

El método DELTA de Burrows indicó que Jorge Joseph es el autor más probable, seguido de Ortega Molina. Este método señala a Emilio Uranga como el menos probable. Este método, al basarse en léxico es sensible a la temática de los textos.

El análisis estilométrico hibrido señaló a Jorge Joseph como principal escritor de ¡El Móndrigo!, seguido de Emilio Uranga y con algunas coincidencias de G. Ortega.

En conclusión, podemos afirmar que los resultados son consistentes: todos señalan a Jorge Joseph como el principal escritor de ¡El Móndrigo! con rastros del estilo de Emilio Uranga.

# Referencias

Burrows, J. (2002). ‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, 267-287.

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

Kilgarriff, A. (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 97-133.

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.

Paass, J. D. (2003). Authorship Attribution with Support Vector Machines. *Applied Intelligence*, 109-123.

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