



## Evidencias lingüísticas del plagio en el periodismo español

Sheila Queralt<sup>1</sup>; Montse Marquina Zarauza<sup>2</sup>; Roser Giménez García<sup>3</sup>

Recibido: 30 de octubre de 2017 / Aceptado: 14 de marzo de 2018

**Resumen.** En disputas legales por posible plagio se requiere la pericia de un lingüista forense. Los estudios en detección de plagio han establecido un umbral máximo del 50% de similitud léxica en textos producidos de forma independiente. En este artículo, se investiga la posibilidad de que los artículos periodísticos requieran un umbral propio puesto que parten de un mismo contenido informativo ("qué", "quién", "cuándo", "dónde", "cómo" y "por qué"). Para ello, se aplican 4 variables lingüísticas cuantitativas a dos corpus estructurados alrededor de 10 temas: un corpus de estudio formado por 50 artículos y un corpus de caso con 20 textos provenientes de un caso real. A partir del primero, se extraen umbrales para cada variable que reflejan los porcentajes de coincidencia esperables entre textos independientes. Estos umbrales se aplican después al corpus del caso para determinar si los nuevos umbrales permiten detectar todos los casos de plagio.

**Palabras clave:** Periodismo; información; plagio; lingüística forense.

### [en] Linguistic evidence of plagiarism in Spanish journalism

**Abstract.** The expertise of a forensic linguist is required in legal disputes about possible plagiarism cases. Studies in plagiarism detection have established a maximum threshold of 50% of lexical similarity in independently produced texts. This paper explores the possibility that journalistic articles require a specific similarity threshold since they share informative content ("what", "who", "when", "where", "how", and "why"). In order to do this, 4 quantitative linguistic variables are applied to two corpora structured around 10 different topics: a corpus of study comprising 50 articles and a case corpus including 20 texts from a real case. On the basis of the former, thresholds for each variable reflecting the expectable coincidence percentages between independent texts are extracted. These thresholds are then applied to the corpus of the case to determine whether the new thresholds allow for all the plagiarism cases to be detected.

**Keywords:** Journalism; information; plagiarism; forensic linguistics..

**Sumario.** 1. Introducción. 2. Marco teórico; 2.1. La lingüística forense; 2.2. Detección de plagio; 2.3. Plagio en el ámbito del periodismo. 3. El caso. 4. Hipótesis y objetivos. 5. Metodología; 5.1. Corpus de estudio; 5.2. Corpus del caso Bryce; 5.3. Análisis; 5.4. Variables. 5.5. *Stop list* del caso. 6.

<sup>1</sup> Laboratorio SQ Lingüistas Forenses (España)  
E-mail: sheila.queralt@cllenciats.cat

<sup>2</sup> Laboratorio SQ Lingüistas Forenses (España)  
E-mail: montserrat.marquina@upf.edu

<sup>3</sup> Laboratorio SQ Lingüistas Forenses (España)  
E-mail: roser.gimenez@cllenciats.cat

Resultados; 6.1. Resultados por tema; 6.1.1. Umbral de similitud; 6.1.2. Frases coincidentes; 6.1.3. Vocabulario exclusivo; 6.1.4. Vocabulario compartido una sola vez; 6.1.5. Resultados finales del corpus de control; 6.2. Resultados del caso; 6.2.1. Umbral de similitud; 6.2.2. Frases coincidentes; 6.2.3. Vocabulario exclusivo; 6.2.4. Vocabulario compartido una sola vez. 7. Discusión. 8. Conclusiones. 9. Referencias bibliográficas.

**Cómo citar:** Queralt, Sheila; Marquina Zarauza, Montse; y Giménez-García, Roser (2018): "Evidencias lingüísticas del plagio en el periodismo español ". *Estudios sobre el Mensaje Periodístico* 24 (2), 1559-1578.

## 1. Introducción

El plagio periodístico ha existido desde hace mucho tiempo, pero hasta hace relativamente poco no se había contemplado como un problema, sino más bien como una realidad unida a la profesión y que, por tanto, suscitaba poco interés.

El desarrollo de las tecnologías de la información y la comunicación (TIC) supuso un arma de doble filo. Por una parte, contribuyó a facilitar el acceso a una gran cantidad de información, pero también contribuyó a que el índice de copias aumentara. En Internet existen millones de documentos que se pueden "reutilizar" de una forma fácil y rápida mediante un simple *copy-paste* ('copiar y pegar'). Este fenómeno ha pasado a ser realmente un problema y objeto de varias investigaciones.

Ramírez Plascencia (2015), por ejemplo, ofrece lúcidas reflexiones sobre la problemática del periodismo digital, tanto por lo concerniente al plagio, como a otras prácticas con posibles consecuencias legales. Según este autor, entre las posibles causas del plagio se encuentran frecuentemente la "ignorancia sobre la forma correcta de citar fuentes y documentos" o "las presiones internas de las empresas para generar el reportaje o la nota a un menor costo y menos tiempo". Además, recuerda con acierto que las consecuencias de esta práctica van más allá del desprestigio de los profesionales responsables, ya que las demandas legales por derechos de autor conllevan costes económicos y manchan la reputación e imagen social de los medios para los que trabajan. Por todas estas razones, es de crucial importancia que los periodistas sean conscientes de que "cualquier acción que haga[n] en internet tendrá forzosamente repercusiones en muchos y muy diversos niveles legales" (2015: 26-27).

## 2. Marco teórico

### 2.1. La lingüística forense

El marco teórico de esta investigación se encuadra dentro de la lingüística forense, disciplina que ocupa el espacio en el que se interrelacionan la lingüística aplicada y el derecho, y utiliza el conocimiento procedente de los ámbitos de las nuevas tecnologías y de la estadística.

La lingüística forense se ha definido, de forma muy general, cubriendo todos los ámbitos en los que se interrelacionan la lengua y el derecho, y, de forma más

restrictiva, como la actuación de los lingüistas en contextos jurídicos y judiciales, a través del uso de pruebas o evidencias lingüísticas.

Así, los peritos lingüistas forenses que trabajan en el ámbito del *lenguaje evidencial* estudian la variación lingüística existente entre los diferentes individuos de una comunidad lingüística determinada, llamada variación interhablante o interescriptor, y la variación presente en un mismo individuo, la variación intrahablante o intraescritor. El principal objetivo del lingüista forense es hallar los rasgos lingüísticos idiosincráticos, único e irrepetibles de los hablantes y de los escritores, para saber si las muestras de texto o de voz que se comparan han sido producidas por la misma persona. Este conjunto de elecciones lingüísticas realizadas por un individuo, que los diferencian del resto de miembros de la comunidad lingüística, es el *idiolecto* (Baldwin, 1979) o *estilo idiolectal* (Turell, 2010). Así pues, se analiza la manera particular en que un individuo utiliza un sistema lingüístico compartido por mucha gente, ya que cada persona usa su lengua de manera distintiva, y este estilo propio es el que es relevante para la lingüística forense.

La presente investigación forma parte del ámbito del lenguaje evidencial, el cual trabaja en dos tipos de evidencia: la evidencia textual y la evidencia oral. Los resultados de este trabajo aportan nuevos datos relevantes para la práctica de la detección de plagio en español.

## 2.2. Detección de plagio

En el ámbito de la detección de plagio, cabe distinguir entre la *copia de ideas* y el *plagio lingüístico*, puesto que mientras que la existencia de copia de ideas sin la necesidad de que haya plagio lingüístico es posible, no puede haber plagio lingüístico sin que se produzca a la vez una copia de ideas.

Hablamos de copia de ideas cuando (Turell, 2011):

- Se utilizan elementos que estructuran una obra literaria: argumento, personajes, tiempo, espacio y técnicas narrativas.
- Se usan todas o casi todas las figuras retóricas de una obra literaria sin citación explícita del autor original, incluso si el vocabulario y la gramática que forman estas figuras son diferentes.
- Se copia una versión traducida cambiando de prosa a verso o transformando una obra clásica en una obra contemporánea, o viceversa.
- Se mencionan los mismos temas para describir un periodo histórico o la contribución de un campo de especialización, en obras científicas.
- Se reproducen los elementos esenciales y la metodología creativa de un texto científico.

Sin embargo, el plagio lingüístico se caracteriza, se gún Turell (2011) —en base a las opiniones de Menasche (1977) y de Roig (2008)—, por las siguientes características:

- Se utilizan las mismas palabras y estructuras gramaticales para describir ideas propias o ajenas.
- Se produce paráfrasis.

- Se usan varias palabras y oraciones sin citar la fuente, aunque se produzcan algunas modificaciones.
- Se mantiene la sintaxis de la obra original y se sustituyen algunas palabras por sinónimos.
- Se reconoce la autoría de la obra original a pesar de que se cambien palabras, el orden de la frase, la voz, el tiempo o el aspecto verbal.

Así pues, el plagio implica tanto la apropiación de una idea como la copia de un texto para expresar dicha idea.

La lingüística forense, en la práctica del análisis o la detección de plagio, se basa en dos principios fundamentales. Por un lado, como plantea Turell (2007), cuando un autor produce un mensaje, sea oral o escrito, crea un texto único e idiosincrático en el que se pueden observar una serie de marcas o recursos lingüísticos que lo hacen irreplicable. En este sentido, Coulthard (2005) afirma que cabe esperar que dos autores que escriben sobre el mismo tema utilicen un conjunto de elementos léxicos y gramaticales que se solapan, pero que en ningún caso serán idénticos. Del otro, en contextos espontáneos, los autores y los receptores no son conscientes de esas marcas lingüísticas y, por lo tanto, pasan inadvertidas para el plagiador que pretende copiarlas o imitarlas (Turell, 2007).

La tarea del perito o lingüista forense es analizar esas marcas o recursos lingüísticos y aportar la mayor cantidad posible de datos lingüísticos para ayudar a confirmar o desmentir la existencia de plagio. Según Turell (2007), la fiabilidad de los datos aportados por los especialistas en el análisis y la detección del plagio depende de la combinación de dos tipos de análisis:

- El análisis cualitativo, basado en criterios lingüísticos como el grado de unidad, de integridad, de coherencia y de cohesión de los textos comparados.
- El análisis cuantitativo, que se basa en una serie de medidas y de métodos analíticos cuantitativos como el nivel de vocabulario coincidente, de palabras compartidas una sola vez o idénticas, y se sirve de herramientas y programas informáticos especializados en la detección de plagio.

El aumento del consumo de las nuevas tecnologías, sobre todo aquellas relacionadas con Internet, ha impulsado la aparición de numerosos estudios, cuantitativos y cualitativos, dedicados a establecer las marcas que permiten identificar el plagio y fijar el umbral de similitud a partir del cual podemos determinar la existencia de plagio, como el trabajo de Turell (2007) o de Coulthard (2006). Sin embargo, todavía se necesitan más investigaciones y recursos para llevarlas a cabo, como un corpus amplio de referencia común.

En relación con el establecimiento de un umbral que permita determinar la existencia de actos de plagio, tomamos como punto de referencia el trabajo realizado por Turell (2007), cuyos resultados, obtenidos mediante varias pruebas empíricas, manifiestan que un umbral de similitud entre dos textos superior al 50% de vocabulario coincidente puede considerarse estándar si ambos textos tratan sobre la misma temática. Por encima de ese valor, incluso con temáticas comunes, se considera que las posibilidades de que uno de los textos haya plagiado al otro son elevadas. Así pues, dicho estudio establece que el umbral de similitud a partir del cual puede considerarse la existencia de plagio es por encima del 50%. Asimismo, se tiene como referencia el estudio realizado por Marquina y Queralt

(2014), en que se establece que es poco probable que en español dos personas produzcan de forma independiente secuencias idénticas de 5 o más palabras.

### 2.3. Plagio en el ámbito del periodismo

Siempre han existido casos de plagio en el mundo del periodismo español, pero hasta hace relativamente poco tiempo no se percibía este acto como negativo o problemático. En los últimos años, sin embargo, se está empezando a criticar y penalizar esta práctica: en el año 2007, la agencia EFE despidió a una de sus redactoras por plagiar un artículo de *Wikipedia*; en 2010, el diario online *Periodista digital* fue condenado por plagiar al bloguero Samuel Parra. En ese mismo año, la periodista Milagros Pérez Oliva publica el artículo “Periodismo de refrito y composición” en el periódico *El País*, advirtiendo del aumento de este tipo de prácticas, sobre todo, a causa de la inmediatez del periodismo digital. Parece que esta tendencia sigue estable actualmente, ya que seguimos encontrando ejemplos de denuncias y críticas del plagio en el periodismo, como recopilaciones de los casos más sonados (p. ej. Hernández Olvera, 2016), iniciativas para detectar el plagio en páginas web (como *churnalism.com*, aunque actualmente se encuentra inactiva por falta de financiamiento) o capítulos sobre esta praxis en libros de ética periodística (p. ej. Jiménez, 2016).

### 3. El caso

La presente investigación se basa en el estudio de un caso: el caso Bryce. El escritor peruano Alfredo Echenique fue denunciado por la Oficina de Derechos de Autor del Instituto Nacional de Defensa de la Competencia y de la Protección a la Propiedad Intelectual (INDECOPI), de Perú, por presunta infracción a los derechos morales de paternidad e integridad, y patrimonial de reproducción de 16 artículos de 15 autores distintos peruanos y españoles. Algunos de los artículos fueron reproducidos casi literalmente (plagio *verbatim*), de periódicos como *La Vanguardia* y revistas como *Jano*.

El autor fue condenado en 2008 a una multa que ascendía a 50 UIT (unos 42 mil euros) en la Resolución No. 142-2008/CDA-INDECOPI de la Comisión de Derecho de Autor de Lima de 24 de diciembre de 2008, la cual fue confirmada en apelación por la Resolución N° 2683-2009-TPI-INDECOPI del Tribunal de Defensa de la Competencia y de la Propiedad Intelectual de Lima de 16 de octubre de 2009.

### 4. Hipótesis y objetivos

Este estudio se plantea un objetivo teórico y dos metodológicos. Por un lado, el objetivo teórico principal de esta investigación es establecer el umbral de similitud a partir del cual puede considerarse que existe plagio entre dos artículos periodísticos, a partir de la observación de un corpus de control y del corpus del caso Bryce.

La hipótesis inicial o hipótesis nula ( $H_0$ ) de la que parte el estudio considera que es posible que el umbral de similitud establecido en los estudios anteriores en el 50% se modifique por la naturaleza del género periodístico.

Por otro lado, el primer objetivo metodológico es la obtención de resultados de investigación derivados de la aplicación de la herramienta analítica *CopyCatch*, y el segundo objetivo metodológico es la aplicación de los resultados de investigación al caso Bryce.

El logro de dichos objetivos nos permitirá poder aportar fiabilidad y validez interna y externa en los resultados en un caso de detección de plagio entre muestras del género periodístico ante un tribunal.

## 5. Metodología

### 5.1. Corpus de estudio

Según el DRAE, la palabra *corpus* designa al “conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”. En este caso, pues, se han recogido 50 artículos periodísticos escritos en español, la mayoría de la variante peninsular, sobre 10 temáticas distintas. Por lo tanto, se han seleccionado 5 artículos de cada tema. Este corpus de control se ha utilizado para establecer el umbral de similitud teniendo en cuenta distintas temáticas.

La selección de temas del corpus de estudio se realizó en base a los de los textos que habían sido plagiados de forma *verbatim* en el caso Bryce, puesto que los resultados ofrecidos por softwares para casos de plagio no *verbatim* son todavía mejorables. Por el momento, son incapaces de reconocer de manera automática varias estrategias empleadas por los plagiarios (como el uso de sinónimos, reformulaciones, reducciones, ampliaciones, etc).

El corpus de estudio se ha realizado y estructurado del siguiente modo (Tabla 1):

Tabla 1. Distribución del corpus de control por tema y media del número de palabras.  
Elaboración propia.

Tema	Artículos (N)	Media de palabras
Tabaco	5	725,4
Grandes potencias mundiales	5	707
Armas nucleares	5	747,8
Terrorismo	5	838,8
Kafka	5	764,8
Guía Michelin	5	819
Elecciones Ségolène Royal	5	958,8
Películas Western	5	509,6
Cultura y civilización	5	1700
Nostalgia	5	1058,6

## 5.2. Corpus del caso Bryce

El corpus perteneciente al caso Bryce está formado por artículos de 10 temáticas diferentes con un número de variables, N, de 2 (el original y el presunto plagiador). Pueden consultarse los textos integrantes en la Tabla 2:

Tabla 2. Distribución del corpus del caso Bryce por comparación tema, título, fecha y autor. Elaboración propia.

Tema	N	Referencia Echenique	Referencia original
Tabaco	2	“Columna incólume: Tabaco y mujer” (01/02/2007)	“Uso social del tabaco” (29/07/2005) Solé, Eulalia
Grandes potencias mundiales	2	“Potencias sin poder”. (12/04/2007)	“Potencias sin poder” Marzo-Abril del 2005. Rivero, Oswaldo de
Armas nucleares	2	“Una amenaza sin fin” (24/05/2007)	“La nueva amenaza nuclear” Mayo del 2005. Rivero, Oswaldo de
Terrorismo	2	“Columna incólume: ¿Cómo combatir el terrorismo?” (01/12/2006)	“¿Cómo combatir el terrorismo?” (29/07 2005). Puigjaner, Josep-Maria
John Ford	2	“John Ford, la épica del <i>western</i> ” (01/07/2006)	“John Ford, la épica del <i>western</i> ” (12/05/2005). Gil Extremera, Blas
Cultura y civilización	2	“Columna incólume: Cultura y civilizaciones” (01 /10/ 2006)	“Cultura y civilizaciones” (27/10/2005). Pera, Cristóbal
Kafka	2	“La angustia de Kafka” (01/10/2003)	“La angustia de Kafka” (25 /10/2001). Ponce, Juan Carlos
Ségolène Royal	2	“Un latido llamado Ségolene” (01/04/2006)	“Ségolene de corazón”, (20/11/2006). Álvaro, Francesc-Marc
Nostalgia	2	“Columna incólume: La enfermedad de la nostalgia” (01/08/2006)	“La enfermedad de la nostalgia” (20/10/2005). Iruela, Luis M.
Estrellas	2	“Estrellas médicas” (01/06/2006)	“Estrellas médicas” (23/04/2004). Pàmies, Sergi

## 5.3. Análisis

El análisis realizado en esta investigación ha consistido en la observación y descripción detallada de los corpus en base a variables cuantitativas mediante el programa informático *CopyCatch Gold*, herramienta de detección de plagio creada por David Woolls en 2002. Dicho análisis describe la manera en que las variables observadas, estudiadas de forma individual, están implicadas en la práctica plagiadora.

## 5.4. Variables

Las principales variables proporcionadas por el programa *CopyCatch* y comúnmente utilizadas por los lingüistas forenses en la comparación de textos para la detección de plagio son cuatro: “umbral de similitud”, “frases coincidentes”, “vocabulario exclusivo” y “vocabulario compartido una sola vez”.

- “Umbral de similitud” (*similarity threshold*) mide el límite de similitud entre los dos textos y lo expresa en porcentajes. Como ya se ha mencionado anteriormente (§ 1.2.), para que se considere la posibilidad de plagio debe haber un grado de similitud superior al 50% entre dos textos.

- “Frasas coincidentes” (*mark up by sentence*) marca el número de frases o sintagmas idénticos o con un alto nivel de similitud entre los textos.

- “Vocabulario exclusivo” (*exclusive vocabulary*) calcula el número de palabras exclusivas de un texto. En este caso, cuanto menor sea el valor de esta variable, más posibilidades hay de que se trate de un plagio.

- Por último, “vocabulario compartido una sola vez” (*shared once*) marca el número de palabras utilizadas una sola vez en cada texto pero que también aparecen en otros textos analizados. Cuanto más elevado sea el valor de esta variable, más similitud habrá entre ambos textos y, por lo tanto, la probabilidad de que se haya plagiado un texto del otro será mayor.

### 5.5. *Stop list* del caso

Con el fin de obtener unos resultados precisos en la comparación del vocabulario coincidente, *CopyCatch* ofrece la posibilidad de incluir una *stop list*, es decir, una lista de palabras que el programa no calculará en el recuento de palabras coincidentes. Esta lista de palabras contiene las palabras funcionales del español (artículos, preposiciones, conjunciones y pronombres), así como los verbos auxiliares o palabras que el lingüista considere que pueden encontrarse en ambos conjuntos dada la temática de los textos.

En este caso, se han incluido en la lista las palabras pertenecientes al campo semántico de los temas tratados en los 50 artículos analizados. Por ejemplo, en el caso del primer tema (“tabaco”), sería imposible hablar de la adicción al tabaco en un artículo sin mencionar las palabras *adicción, ayuda, cigarrillo, dejar, fumar, fumador, nicotina, tabaco, tabaquera*.

## 6. Resultados

Los resultados extraídos del corpus de control (y que, por tanto, no pueden reflejar plagio entre los textos) mediante el programa *CopyCatch* se representan a continuación en forma de gráficos. En primer lugar, se ha analizado cada uno de los temas por separado y, posteriormente, se ha realizado una comparación global agrupando los datos proporcionados por cada tema.

En la Tabla 3 se puede observar un extracto de la comparativa arrojada por el programa *CopyCatch* mediante la opción “frases coincidentes” (*Mark up by sentence*) del tema “tabaco”, entre dos textos del corpus de control. Esta tabla muestra de forma visual secuencias de palabras coincidentes entre dos textos. En color rojo se indican las coincidencias y en color negro las diferencias. En el caso de encontrar similitudes, el programa ofrece el número del párrafo y la oración (“[P6 S1]”) con su correspondencia en el otro texto, en este caso coincidente (“{P12 S2}”).



Tabla 3. Extracción del resultado comparativo del tema “tabaco” del corpus de estudio mediante el programa *CopyCatch* (variable “frases coincidentes”). Elaboración propia.

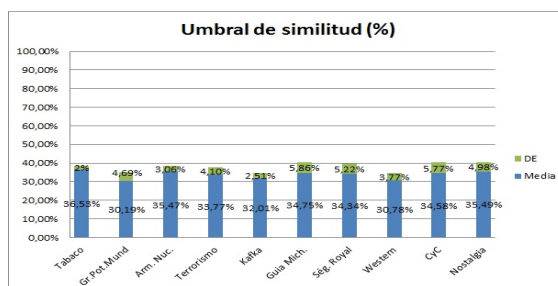
Tabaco 04	Tabaco 05
<p>[P6 S1] <i>¿Sirven para dejar de fumar tabaco?</i> {P12 S2} [P6 S2] Buena parte de los fumadores han pensado que quizás. [P6 S3] La experiencia de los médicos que ayudan a dejar el tabaco es que una parte de los que lo han probado acaban abandonando (el electrónico) porque fumaban doble. [P6 S4] Otros han conseguido reducir la cantidad de cigarrillos. [P6 S5] Un estudio publicado en The Lancet este verano incluso defendió que podría ser tan útil como el parche de nicotina. [P6 S6] Y los médicos deshabitadores no descartan su uso para fumadores extremos (como entre los enfermos mentales) cuando haya alguna seguridad. [P6 S7] Pero hay dos problemas (además de no saber qué te metes). [P6 S8] Uno, que la nicotina llega en segundos al cerebro, como el tabaco, mientras que con el parche o los chicles tarda un par de horas y eso ayuda a reducir la dependencia. [P6 S9] Dos, que se parece tanto a fumar, que el ritual -el problema más difícil para dejar el tabaco en la mayoría- no se resuelve en absoluto. [P6 S10] Al contrario, se perpetúa. [P6 S11] "Es mejor llevarlo colgado, así le das una calada de vez en cuando", sugiere la vendedora.</p>	<p>{P12 S1} "No podemos decir que un fabricante de cigarrillos electrónicos use los mismos componentes que otro", señala Erika Sward, de la Asociación Pulmonar Estadounidense. {P12 S2} <b>Aunque se clasifican como productos para dejar de fumar, algunos profesionales sanitarios advierten de su toxicidad.</b> [P6 S1] {P12 S3} "Atención, consumidores", señala.</p>

### 6.1. Resultados por tema

A continuación, se muestran los resultados procedentes de cada uno de los temas del corpus de control. Los resultados se desglosan en las 4 variables propuestas: “umbral de similitud”, “frases coincidentes”, “vocabulario exclusivo” y “vocabulario compartido una sola vez”.

#### 6.1.1. Umbral de similitud

Gráfico 1. Media de similitud y desviación típica en cada uno de los temas. Elaboración propia.

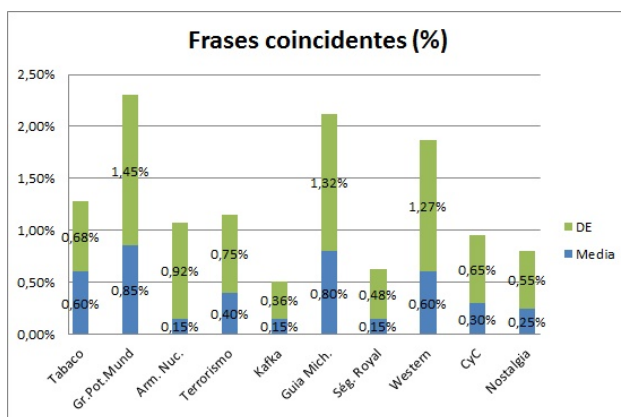


En este gráfico de barras se puede observar la media de similitud y la desviación típica en cada uno de los temas. La media más alta se sitúa en un 36,52% en el tema “tabaco” y la media más baja en un 30,19% en el tema “grandes potencias mundiales”. Si se tiene en cuenta la desviación típica, se puede observar que el porcentaje más elevado se encuentra en el tema “Guía Michelin”, con un 40,61%, y el más bajo, con un 34,51%, en el tema “Kafka”.

Estos porcentajes son inferiores al 50% de similitud establecido en estudios anteriores, lo que indicaría que el umbral de similitud necesario para detectar plagio entre artículos periodísticos no excedería el umbral estándar e incluso se podría marcar en un nivel inferior.

### 6.1.2. Frases coincidentes

Gráfico 2. Frases coincidentes (%). Elaboración propia.



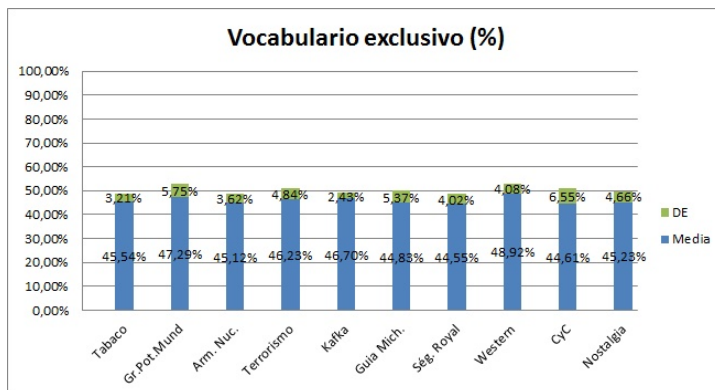
En este segundo gráfico de barras observamos las frases coincidentes marcadas en el corpus de control. Dada la escala del gráfico, a simple vista pueden parecer resultados bastante diferentes para cada tema. Sin embargo, a una escala mayor, podemos observar que todos muestran porcentajes muy bajos y, por tanto, parecidos.

La media más alta que podemos observar corresponde al tema “grandes potencias mundiales”, con un 0,85%, y la media más baja se sitúa en los temas “armas nucleares”, “Kafka” y “Ségolène Royal”, con un 0,15%. Si nos fijamos en la desviación típica, encontramos que el porcentaje más alto corresponde al tema “grandes potencias mundiales” con un porcentaje total de 2,3%. El porcentaje más bajo se encuentra en el tema “Kafka”, con un 0,51%.

Estos datos muestran el porcentaje de frases idénticas en los diferentes temas tratados, lo cual hace posible establecer un umbral de sospecha de plagio (concretamente, en el caso de esta variable, de plagio *verbatim*). El establecimiento del umbral se realiza en la sección 6.1.5. mediante el Gráfico 5.

### 6.1.3. Vocabulario exclusivo

Gráfico 3. Vocabulario exclusivo (%). Elaboración propia.



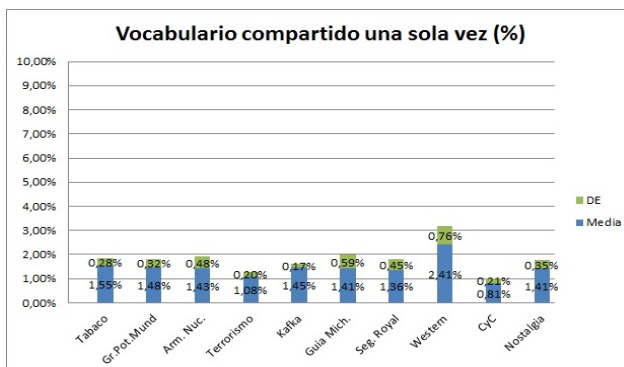
En este diagrama de barras podemos observar la variable “vocabulario exclusivo” de los diferentes temas recogidos en el corpus de control. El gráfico muestra valores muy elevados que pueden sugerir otro umbral a partir del cual deberíamos sospechar que existe plagio. Los resultados para esta variable muestran una uniformidad considerable.

El porcentaje más elevado, si nos fijamos en la media, se encuentra en el tema “cultura y civilización”, con un 48,92%, y el porcentaje más bajo, en el tema “Guía Michelin”, con un 44,83%. Si contemplamos la desviación típica, situamos el porcentaje más alto en un 53,04% para el tema “grandes potencias mundiales”, y el más bajo en un 48,57% para “Ségolène Royale”.

Estos porcentajes permiten intuir el umbral que se podrá establecer para esta variable, especialmente para poder establecer la existencia de plagio mediante ampliación, así como mediante reducción (ver apartado 6.1.5. y Gráfico 5).

### 6.1.4. Vocabulario compartido una sola vez

Gráfico 4. Vocabulario compartido una sola vez (%). Elaboración propia.



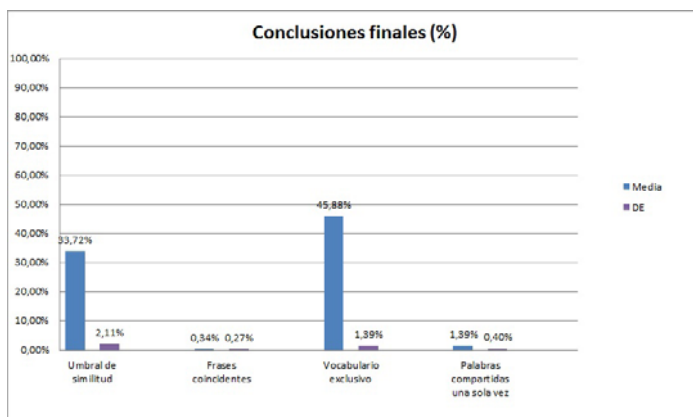
Este diagrama de barras nos muestra la variable del vocabulario compartido una sola vez en los varios temas del corpus de control. Se trata de porcentajes bastante bajos. Podemos observar una similitud en los resultados entre los temas, lo que hace posible pensar en el establecimiento de un umbral concreto para esta variable.

El porcentaje representado más alto, teniendo en cuenta la media, corresponde al tema “Western”, con un 2,41%, y el porcentaje más bajo corresponde al tema “cultura y civilización”, con un 0,81%. Si observamos la desviación típica, encontramos el porcentaje más elevado también en el tema “Western” con un 3,17% y el porcentaje más bajo en el tema de “cultura y civilización” con un 1,02%. Así pues, ambas medidas coinciden en esta variable para los temas con valores mayores y menores.

El establecimiento del umbral se realizará en el punto 6.1.5. con la ayuda del Gráfico 5.

### 6.1.5. Resultados finales del corpus de control

Gráfico 5. Conclusiones finales (%). Elaboración propia.



En este diagrama de barras podemos observar las medias finales obtenidas para las diferentes variables del corpus de control con sus respectivas desviaciones típicas. Estos resultados permiten establecer un umbral para sospechar la ocurrencia de plagio según cada variable en el caso de artículos periodísticos.

En la variable “umbral de similitud”, podemos observar una media de 33,72% con una desviación típica del 2,11%, lo cual proporciona una media final de 35,83%. Así pues, estableceríamos el umbral para la variable “umbral de similitud” en un 33,72%, pudiendo alargar este margen hasta un 35,83%. Este porcentaje es considerablemente más bajo que el 50% establecido para otros estudios, lo cual confirmaría la modificación del porcentaje de coincidencia necesario en esta variable para la aparición de sospecha de plagio en artículos periodísticos.

La variable “frases coincidentes” presenta una media de 0,34% con una desviación típica del 0,27%. Por lo tanto, partiendo de la media, el margen de sospecha para esta variable sería del 0,61%, lo cual representa el margen de sospecha de la existencia de plagio *verbatim*.

La variable “vocabulario exclusivo” muestra una media del 45,88% con una desviación típica del 1,39%. Teniendo en cuenta la naturaleza de la variable, aplicaremos la desviación típica en el sentido contrario a las demás para poder establecer el umbral a partir del cual podría tratarse de plagio. Así pues, aplicando el valor de la desviación típica a la media, obtendríamos un umbral en un 45,88% teniendo en cuenta la desviación típica.

Finalmente, la variable “vocabulario compartido una sola vez” muestra un porcentaje del 1,39% con una desviación típica del 0,40%. Esto nos permite establecer en este porcentaje (1,39%) el umbral en el cual sospechar de plagio. Si tomamos en consideración el margen que nos proporciona la desviación típica, podemos situar el porcentaje máximo de coincidencia en un 1,79%.

## 6.2. Resultados del caso

Los resultados para el caso “Echenique” se han extraído utilizando el programa informático *CopyCatch Gold* mediante la comparación de dos variables en cada caso: el texto del autor Bryce Echenique (presuntamente plagiado) y el texto presuntamente original del cual Echenique habría realizado el plagio.

En la Tabla 4 se puede observar la comparativa arrojada por el programa *CopyCatch* del tema “tabaco” entre el texto original de la autora Eulàlia Solé y el texto de Bryce. A diferencia de lo que ocurría entre la comparativa de los textos de control (ver Tabla 3), se puede observar que el color rojo destaca puesto que el plagio *verbatim* entre los textos es total. Las palabras de color negro (y, por tanto, las diferencias) son mínimos y corresponden a las incorporaciones o cambios propios de las estrategias de un plagiario: sinónimos, reducciones, ampliaciones, modificaciones de estilo (como la inclusión de las comillas), etc.

Tabla 4. Extracción del resultado comparativo del tema “tabaco” del corpus del caso mediante el programa *CopyCatch* (variable “frases coincidentes”). Elaboración propia.

“Uso social del tabaco” (29/07/2005) de Solé, Eulalia	“Columna incólume: Tabaco y mujer” (01/02/2007) de Bryce
<p><b>[P1 S1] USO SOCIAL DEL TABACO</b>  <b>[P2 S1]</b> Según un estudio publicado por la revista <i>Addiction</i>, las grandes empresas tabaqueras son responsables, en buena parte, de que el hábito de fumar se haya incrementado entre las mujeres. <b>{P2 S1}</b> <b>[P2 S2]</b> Cajetillas con modalidades como light, ligero, o slim, esbelto, pretenden asociar el consumo a atributos como libertad, esbeltez, glamour, y lo cierto es que actualmente las mujeres fuman tanto o más que los hombres. <b>{P2 S2}</b> <b>[P2 S3]</b> ¿Cuál ha sido el camino que ha conducido hasta ese punto desde la introducción del tabaco en Occidente? <b>{P3 S1}</b>  <b>[P3 S1]</b> Primero fueron los hombres, y además, ricos. <b>{P4 S1}</b> <b>[P3 S2]</b> Cuando los conquistadores españoles descubrieron que los indios americanos consumían esta hierba, y no sólo en forma de canutillos encendidos cuyo humo aspiraban y exhalaban, sino comiéndola y también convirtiéndola en bebida, pronto exportaron el producto a la metrópoli. <b>{P4 S2}</b> <b>[P3 S3]</b> Aunque fue sólo en la vertiente de fumar, pronto se convirtió en un</p>	<p><b>{P1 S1} TABACO Y MUJER</b>  <b>{P2 S1}</b> Según un estudio publicado por la revista <i>Addiction</i>, las grandes empresas tabaqueras son responsables en buena parte de que el hábito de fumar se haya incrementado entre las mujeres. <b>[P2 S1]</b> <b>{P2 S2}</b> Cajetillas con modalidades como light, slim, esbelto, o ligero, pretenden asociar el consumo a atributos como libertad, esbeltez, glamour, y lo cierto es que actualmente las mujeres fuman tanto o más que los hombres. <b>[P2 S2]</b>  <b>{P3 S1}</b> ¿Cuál ha sido el camino que ha conducido hasta ese punto desde la introducción del tabaco en Occidente? <b>[P2 S3]</b>  <b>{P4 S1}</b> Primero fueron los hombres, y además, ricos. <b>[P3 S1]</b> <b>{P4 S2}</b> Cuando los conquistadores españoles descubrieron que los indios americanos consumían esta hierba, y no sólo en forma de canutillos encendidos cuyo humo aspiraban y exhalaban, sino comiéndola y también convirtiéndola en bebida, pronto exportaron el producto a la metrópoli. <b>[P3 S2]</b> <b>{P4 S3}</b> Aunque fue</p>

gran negocio, de tal manera que el Reino de España adjudicó a Hacienda, el año 1623, el monopolio de la venta de tabaco. {P4 S3} {P3 S4} Y así ha seguido hasta hace poco. {P4 S4} {P3 S5} Las grandes plantaciones estaban en Cuba y Santo Domingo, y pronto el hábito se extendió a las clases populares a tenor del aumento de las importaciones y el descenso del precio. {P4 S5}

{P4 S1} Hay que advertir de que no en todos los países se dio al tabaco la misma acogida. {P5 S1} {P4 S2} En muchos lugares fue tachado de droga detestable, por el humo que provocaba, así como por el mal olor que despedía. {P5 S2} {P4 S3} Durante el siglo XVII, tanto la venta como el consumo estaban castigados en Europa central, Rusia, Turquía y China, naciones que, curiosamente, no participaban en el ejercicio de la exportación/importación. {P5 S3} {P4 S4} Nada impidió, no obstante, que el tabaco acabara por instaurarse en las costumbres sociales tanto occidentales como orientales, reservado de entrada al género masculino. {P5 S4}

{P5 S1} El gesto de fumar equivalía, para los hombres, a un signo de virilidad, mientras que las pocas mujeres que se atrevían a fumar eran unas transgresoras que, según las miradas, resultaban interesantes o desvergonzadas. {P6 S1}

{P6 S1} Lo que nadie sabía, todavía, es que los fumadores estaban jugando con una droga adictiva y peligrosa para la salud. {P7 S1} {P6 S2} Se fumaba en casa, en la oficina, en el cine y el teatro, en los transportes públicos, incluso en los hospitales. {P7 S2} {P6 S3} No sólo en los pasillos sino en las propias habitaciones, por parte de médicos y enfermos. {P7 S4} {P7 S1} Lo sorprendente, tanto como inquietante, es que el descubrimiento de los atributos nocivos del tabaco no haya redundado en el descenso tajante de su consumo. {P8 S1} {P7 S2} La explicación ha de hallarse justamente en esos atributos, los de causar dependencia psíquica y física, amén de desarrollar una tolerancia que incita a fumar cada vez más para sentirse igual de estimulado. {P8 S2} {P7 S3} Si a esto agregamos la publicidad, incluso la engañosa para las mujeres citada más arriba, nada tiene de extraño que chicos y chicas se enganchen a edades progresivamente tempranas. {P8 S3}

{P8 S1} En cuanto a las mujeres, adultas, diversos factores explican su masiva incorporación al consumo del tabaco. {P9 S1} {P8 S2} Emancipadas, integradas en el mundo laboral masculino, el tabaco se ha convertido, tal como siglos atrás sucedió con ellos, en un símbolo. {P9 S2} {P8 S3} Por desgracia, el gesto de fumar ya forma parte de la feminidad. {P9 S3}

sólo en la vertiente de fumar, rápidamente se convirtió en un gran negocio, de tal manera que el Reino de España adjudicó a Hacienda, en 1623, el monopolio del tabaco. {P3 S3} {P4 S4} Y así ha seguido hasta hace poco. {P3 S4} {P4 S5} Las grandes plantaciones estaban en Cuba y Santo Domingo, y pronto el hábito se extendió a las clases populares a tenor del aumento de las importaciones y el descenso de precio. {P3 S5}

{P5 S1} Hay que advertir de que no en todos los países se dio al tabaco la misma acogida. {P4 S1} {P5 S2} En muchos lugares fue tachado de droga detestable, por el humo que provocaba, así como por el mal olor que despedía. {P4 S2} {P5 S3} Durante el siglo XVII, tanto la venta como el consumo estaban castigados en Europa central, Rusia, Turquía y China, naciones que curiosamente no participaban en el ejercicio de la exportación/importación. {P4 S3} {P5 S4} Nada impidió, no obstante, que el tabaco acabara por instaurarse en las costumbres sociales tanto occidentales como orientales, reservado de entrada al género masculino. {P4 S4}

{P6 S1} El gesto de fumar equivalía, para los hombres, a un signo de virilidad, mientras que las pocas mujeres que se atrevían a fumar eran unas transgresoras que, según las miradas, resultaban interesantes, desvergonzadas o machonas. {P5 S1}

{P7 S1} Lo que nadie sabía todavía es que los fumadores estaban jugando con una droga adictiva y muy peligrosa para la salud. {P6 S1} {P7 S2} Se fumaba en casa, en la oficina, en el cine y el teatro, en los transportes públicos e incluso en los hospitales. {P6 S2} {P7 S3} Y no sólo en los corredores sino en las propias habitaciones. {P7 S4}

Y fumaban médicos y enfermos. {P6 S3}

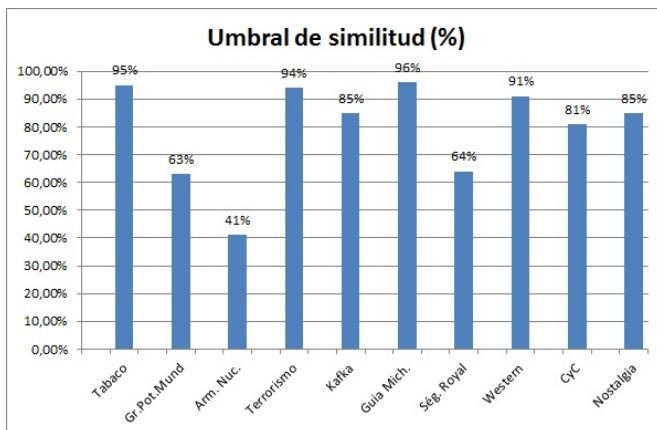
{P8 S1} Lo sorprendente, tanto como inquietante, es que el descubrimiento de los efectos nocivos del tabaco no haya contribuido al descenso tajante de su consumo.

{P7 S1} {P8 S2} La explicación debe hallarse justamente en esos atributos, los de causar dependencia física y psíquica, y a la vez desarrollar una tolerancia que incita a fumar cada vez más para sentirse igual de estimulado. {P7 S2} {P8 S3} Si a esto agregamos la publicidad, incluso la engañosa para las mujeres, citada más arriba, nada tiene de extraño que chicos y chicas se “enganchen” hoy a edades progresivamente tempranas. {P7 S3}

{P9 S1} En cuanto a las mujeres adultas, diversos factores explican su masiva adicción al consumo del tabaco. {P8 S1} {P9 S2} Emancipadas, integradas en el mundo laboral masculino, el tabaco se ha convertido, tal como siglos atrás sucedió con ellos, en un símbolo. {P8 S2} {P9 S3} Por desgracia el gesto de fumar ya forma parte de la feminidad. {P8 S3}

### 6.2.1. Umbral de similitud

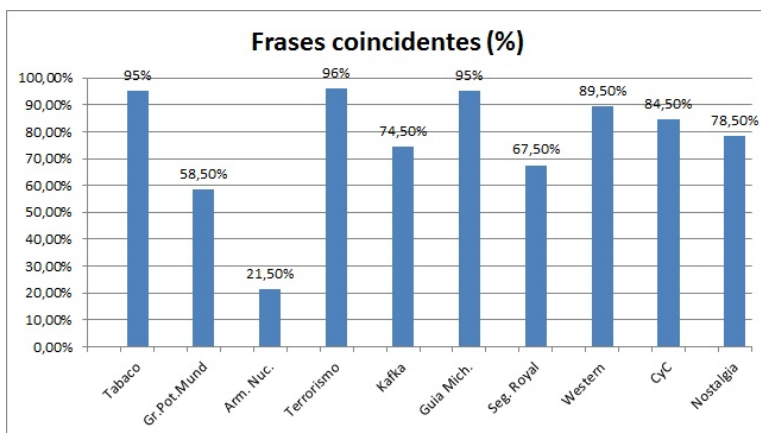
Gráfico 6. Umbral de similitud (%). Elaboración propia.



En este diagrama de barras podemos observar los varios valores que toma la variable “umbral de similitud” en los temas observados en el corpus del caso real. Detectamos una tendencia a representar porcentajes muy elevados. Si aplicáramos el porcentaje proporcionado por estudios anteriores (50%), podríamos afirmar la sospecha de plagio en todos los temas excepto en “armas nucleares”. En cambio, si aplicamos el umbral establecido en este estudio para artículos periodísticos 35,83%, podemos establecer la sospecha de plagio para todos los temas mostrados en el caso real.

### 6.2.2. Frases coincidentes

Gráfico 7. Frases coincidentes (%). Elaboración propia.



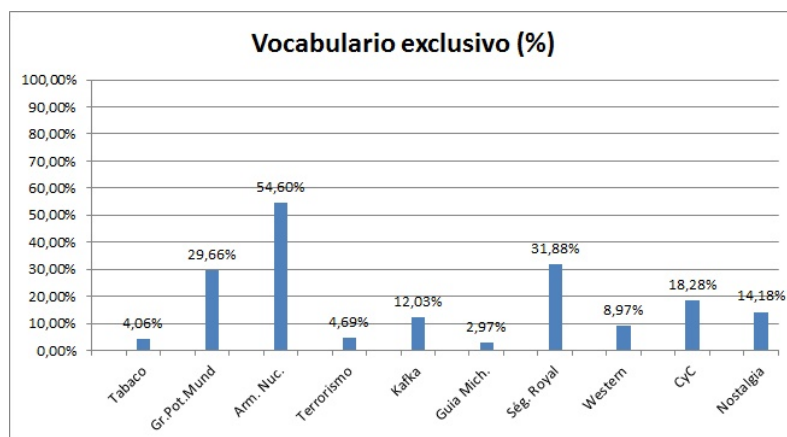
Este diagrama de barras nos muestra los valores resultantes para la variable “frases coincidentes” para cada uno de los temas comprendidos en el corpus del caso real. Así pues, podemos observar el valor más alto en el tema “terrorismo” con un valor

del 96% y el valor más bajo en “armas nucleares”, que muestra un valor de 21,50%.

Siguiendo los resultados obtenidos por estudios anteriores, la variable “armas nucleares” se consideraría dudosa de ser representar plagio, mientras que todas las demás se considerarían plagios sistemáticamente. Usando el umbral establecido (0,61%) en este trabajo, consideraríamos que todos los temas han sido plagiados de forma flagrante.

### 6.2.3. Vocabulario exclusivo

Gráfico 8. Vocabulario exclusivo (%). Elaboración propia.



En este diagrama de barras se muestran los valores obtenidos para la variable “vocabulario exclusivo”. El tema que muestra un valor más elevado corresponde al tema “armas nucleares” con un 54,60%. Por otro lado, vemos que el valor más bajo corresponde al tema “Guía Michelin” con un 2,97%.

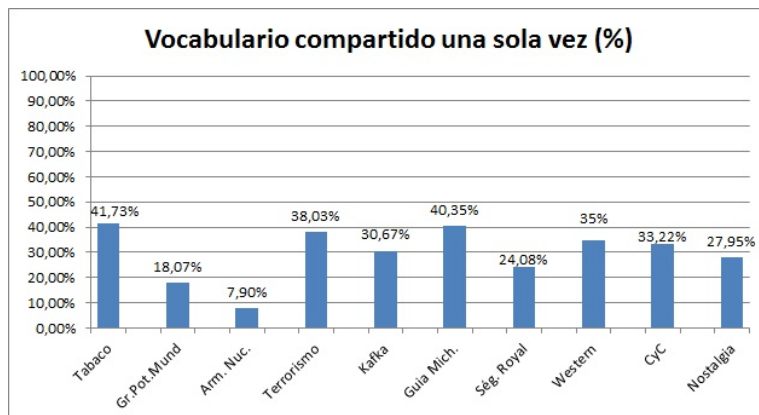
Teniendo en cuenta los estudios anteriores podríamos concluir que no se puede constatar plagio en los temas “armas nucleares” y, en menor medida, “Ségolène Royal” y “grandes potencias mundiales”. Aun así, si aplicamos el umbral obtenido anteriormente de un 44,49% tanto en el tema “Ségolène Royal” como “grandes potencias mundiales”, los resultados apoyan la sospecha de haberse cometido plagio. En cambio, el tema “armas nucleares” continuaría fuera de sospecha después de aplicar el umbral propio.

### 6.2.4. Vocabulario compartido una sola vez

El diagrama de barras del siguiente gráfico (9) muestra los resultados obtenidos en el caso real para la variable “vocabulario compartido una sola vez” divididos por temas y representados en porcentajes. Como se puede observar, el tema con un porcentaje mayor corresponde a “tabaco”, con un 41,73%, mientras que el tema con un porcentaje menor corresponde a “armas nucleares”, con un 7,90%. En todos los casos se supera el porcentaje máximo de coincidencia esperable (1,79%).



Gráfico 9. Vocabulario compartido una sola vez (%). Elaboración propia.



## 7. Discusión

El principal objetivo teórico de esta investigación era establecer el umbral de similitud entre artículos periodísticos que sugiera que su producción no ha sido independiente, y se podría estar incurriendo en un delito contra la propiedad intelectual de un tercero. Este objetivo se ha alcanzado mediante el logro de los objetivos metodológicos, la aplicación de la herramienta de comparación forense de textos *CopyCatch* en dos corpus comparables: el corpus de control y el corpus compuesto por textos del caso de plagio real. La observación de 4 variables distintas permite contrastar los resultados para cada una y llegar a una conclusión firme sobre la hipótesis inicial de que exista una diferencia entre el umbral de similitud identificado en estudios previos y el necesario para el análisis en textos periodísticos.

Dicha hipótesis se ha confirmado claramente en el caso de la primera variable, “umbral de similitud”, ya que, mientras la aplicación del umbral preestablecido del 50% de coincidencia dejaba fuera de sospecha a uno de los textos del caso, el uso del umbral extraído de la comparación de los textos del corpus de control (35,83%) permitió sospechar de plagio en todos los artículos del caso Bryce. Asimismo, en el caso de la cuarta variable, “vocabulario compartido una sola vez”, todos los textos del caso Bryce producen resultados que sugieren que fueron plagiados según el umbral establecido con el estudio del corpus de control. Por eso, es posible afirmar que el umbral específico para el género periodístico produjo unos resultados que se ajustan mejor a los deseables (es decir, la detección de plagio en todos los casos en que se da) que el umbral de similitud establecido en otros estudios.

De forma parecida, para la segunda variable, que calcula el plagio *verbatim* mediante el recuento de frases idénticas, uno de los temas del corpus del caso proporciona resultados dudosos con la aplicación del umbral establecido anteriormente. Aun así, el uso del nuevo umbral específico para el género

periodístico, establecido a través de los resultados del corpus de control (0,61%), permite identificar plagio en todos los casos.

Algo diferente ocurre con la tercera variable, “vocabulario exclusivo”, ya que solo dos de los tres temas excluidos de la sospecha de plagio mediante el umbral preestablecido se confirman como posibles plagios al aplicar el umbral obtenido con el corpus de control (44,49%).

Mediante la aplicación sistemática de metodología estándar en la práctica forense de comparación de textos para la detección de plagio, en esta investigación se han obtenido resultados para 4 variables. Turell (2011: 76) recuerda la necesidad en casos forenses reales de utilizar todos los recursos proporcionados por este programa, ya que ninguna de las 4 variables posee una capacidad discriminatoria suficiente como para obtener resultados concluyentes sin las demás. De este modo, los umbrales obtenidos en este estudio podrán ser aplicados en futuros casos e investigaciones de plagio basados en textos periodísticos. Además, se ha demostrado una metodología que permite investigar si los umbrales generales establecidos en estudios anteriores pueden ser ajustados para otras tipologías textuales específicas. La existencia de estudios como el presente es clave para la práctica de la lingüística forense y la demostración de la fiabilidad interna y externa en su uso en casos judiciales.

## 8. Conclusiones

En base a los resultados obtenidos en esta investigación podemos afirmar que el umbral de similitud necesario para determinar la existencia de plagio entre dos artículos periodísticos es inferior al 50% establecido para la detección de plagio de textos en general. En el caso específico de los textos pertenecientes al género periodístico, el umbral se sitúa entre el 35% y el 40% de similitud. Así pues, los resultados obtenidos del análisis realizado en este estudio nos permiten aceptar la hipótesis inicial de la que parte esta investigación (§ 2).

Como se ha demostrado, el uso del umbral de similitud específico para la detección de plagio en artículos periodísticos puede aumentar el éxito de identificación de casos de plagio. En esta investigación, por ejemplo, algunos textos incluidos en el tema “armas nucleares” presentaron un grado de similitud menor al 50%, por lo que, con el umbral de similitud general, no se considerarían instancias de plagio. No obstante, dado que su similitud supera el umbral del 40% establecido por este estudio para artículos periodísticos, sí deberían considerarse sospechosos de constituir plagio.

Parece, pues, que los artículos periodísticos poseen características propias que los diferencian de los textos pertenecientes a otros géneros y, por lo tanto, deben ser examinados desde otra óptica en la práctica y detección de plagio.

## 9. Referencias bibliográficas

- Baldwin, John R. (1979): "Phonetics and speaker identification". *Medicine, Science and the Law*, 19 (4), 231-232.
- Coulthard, Malcolm (2005): "The linguist as expert witness". *Linguistics and the Human Sciences*, 1 (1), 39-58.
- Coulthard, Malcolm (2006): *...And then... language description and author attribution*. En: <http://www.aston.ac.uk/EasySiteWeb/GatewayLink.aspx?allId=10186> [Consulta: 26 de octubre de 2017]
- Hernández Olvera, Katheryn (2016): "Diez casos de plagios periodísticos que no debes olvidar". *Clases de periodismo*, 27 de enero. En: <http://www.clasesdeperiodismo.com/2016/01/27/diez-casos-de-plagios-periodisticos-que-no-debes-olvidar/> [Consulta: 26 de octubre de 2017]
- Jiménez, Roger (2016): *Cien casos. La ética periodística en tiempos de precariedad*. Barcelona, Edicions Universitat Barcelona.
- Marquina, Montse y Queralt, Sheila (2014): "Similarity threshold to detect plagiarism in Spanish". *RaeL: Revista Electrónica de Lingüística Aplicada*, 13 (1), 79-95.
- Menasche, Lionel (1977): *Writing a Research Paper*. Ann Arbor, The University of Michigan Press.
- Ramírez Plascencia, David (2015): "Las disyuntivas legales del periodismo en la era digital". *Acta Republicana. Política y Sociedad*, 14, 23-28.
- Real Academia Española (2001): *Diccionario de la lengua española* (22ª ed.). En: [www.rae.es/rae.html](http://www.rae.es/rae.html) [Consulta: 26 de octubre de 2017]
- Roig, Miguel (2008): *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*. En: <https://ori.hhs.gov/images/ddblock/plagiarism.pdf> [Consulta: 26 de octubre de 2017]
- Turell, Maria Teresa (2007): "Textual kidnapping revisited: The case of plagiarism in literary translation". *International Journal of Speech Language and the Law*, 11 (1), 1-26.
- Turell, Maria Teresa (2010): "The use of textual, grammatical and sociolinguistic evidence in forensic text comparison". *The International Journal of Speech, Language and the Law*, 17 (2), 211-250.
- Turell, Maria Teresa (2011): "La tasca del lingüista detectiu en casos de detecció de plagi i determinació d'autoria de textos escrits". *Llengua, societat i comunicació*, (9), 69-85.

---

Sheila Queralt, doctora en Traducción y Ciencias del Lenguaje por la Universitat Pompeu Fabra (UPF), es directora del Laboratorio SQ-Lingüistas Forenses. Es docente de máster en la Facultad de Comunicación y Ciencias Sociales de la Universidad San Jorge, docente en el Centro de Investigación y Formación Criminológica de la Universidad Miguel Hernández y directora académica de los cursos profesionales en Lingüística Forense del Laboratorio SQ – Lingüistas Forenses. Es perito colegiada en el Col·legi Oficial de Doctors i Llicenciats en Filosofia i Lletres i Ciències como perito judicial en lingüística forense y en criminalística de la Propiedad Intelectual e Industrial.

Montse Marquina Zarauza, doctora en Traducción y Ciencias del Lenguaje por la Universitat Pompeu Fabra (UPF), es perito sénior en lingüística forense en la especialidad de análisis del habla y de textos escritos en el Laboratorio SQ-Lingüistas Forenses. Ha trabajado como investigadora en el Departamento de Lengua Española y Lingüística General de la UNED, en el Grupo de Fonética de la UAB, en el grupo de Voz y Lenguaje de Barcelona Media y en el ForensicLab-UVAL (Unitat de Variació Lingüística) del Institut de Lingüística Aplicada (Centre específic de recerca) de la Universitat Pompeu Fabra.

Roser Giménez es perito e investigadora en lingüística forense en comparación forense de textos escritos en el Laboratorio SQ – Lingüistas Forenses. Es perito colegiada en el Col·legi Oficial de Doctors i Llicenciats en Filosofia i Lletres i Ciències como perito judicial en lingüística forense. Es máster en lingüística forense por Aston University en el Reino Unido. Es miembro del equipo docente en los cursos profesionales del Laboratorio SQ – Lingüistas Forenses.