ANÁLISIS Y HERRAMIENTAS PARA LA DETECCIÓN DE PLAGIO APLICADAS A LA TESIS DE PEDRO SÁNCHEZ

Daniel Guirado Aguilar

(Universidad de Murcia. Egresado del Máster universitario en Lingüística Teórica y Aplicada. Almería, España)

dani.guirado.aguilar@gmail.com

Juan Antonio Cutillas Espinosa

(Universidad de Murcia. Facultad de Letras. Departamento de Filología Inglesa. Murcia, España)

jacuti@um.es

ANALYSIS AND TOOLS FOR PLAGIARISM DETECTION APPLIED TO THE PHD DISSERTATION WRITTEN BY PEDRO SÁNCHEZ

Fecha de recepción: 7-12-2019 / Fecha de aceptación: 16.06.2020

Tonos Digital, 39, 2020 (II)

RESUMEN

El plagio es un problema que afecta tanto a la comunidad científica como la artística y, a pesar de que la Lingüística Forense trata de avanzar cada vez más en su investigación, sigue siendo un campo donde todavía hay pocos estudios públicos sobre ello, en tanto que ni siquiera hay un método científico convencionalizado para su detección.

Es por todo esto por lo que nosotros hemos tratado de profundizar en esta materia, llevando a cabo, en primer lugar, una revisión sobre el plagio, con la que describiremos sus principales características, así como los métodos que existen para su detección, con el fin de analizar posteriormente desde el punto de vista académico la tesis de Pedro Sánchez Castejón, el actual presidente del Gobierno de España, ya que ha sido ampliamente criticada recientemente por los medios de comunicación, así

como también ha servido como arma política, al calificarla completamente como plagio, sin aparentemente conocer las implicaciones que ello conlleva.

Luego, tomaremos una muestra de las referencias indicadas por Turnitin como coincidentes, con el fin de calcular la precisión de esta herramienta para detectar el plagio, utilizando para ello una categorización propia que nos permitirá conocer, además, las limitaciones del programa antiplagio, tal vez, más usado del mundo.

Finalmente, se analizará exhaustivamente el resto de las referencias del listado de esta herramienta para determinar si ha existido algún tipo de presunto plagio en la tesis de Pedro Sánchez, indicando el porcentaje exacto de presunto plagio y si ha podido ser o no intencional.

Palabras clave: Plagio; métodos de detección de plagio; antiplagio; Lingüística Forense; tesis de Pedro Sánchez Castejón.

ABSTRACT:

Plagiarism is a problem that affects both the scientific and the artistic community, and, despite the fact that Forensic Linguistics keeps on investigating it, there are so few studies about plagiarism that there is no conventionalised method to deal with its detection.

It is for this reason that our investigation tries to delve into this subject, carrying out a review of the concept of plagiarism. It includes an academic and legal description of its main features, as well as a discussion of plagiarism detection methods. We then focus on the case study of the doctoral thesis written by the current Spanish President, Pedro Sánchez Castejón. This dissertation was the object of recent plagiarism accusations in news media, which some opposition representatives used as a political weapon, describing the thesis as a whole as the result of plagiarism. This generalization reveals a lack of insight into the complexities of this concept.

We will take a sample of some of the matches between the thesis and the list of references provided by Turnitin, as a means to measure this tool's accuracy in detecting plagiarism. We will use our own categorization, which will enable us to explore the limitations of this antiplagiarism software.

Finally, we will analyse the rest of the references thoroughly to determine whether the thesis contains some instances of actual plagiarism, pointing out the exact percentage of it and whether it is intentional or unintentional.

Keywords: plagiarism; plagiarism detection methods; antiplagiarism; Forensic Linguistics; Pedro Sánchez Castejón's doctoral thesis.

1. INTRODUCCIÓN.

En la actualidad, tanto la comunidad científica como la artística corren el riesgo de que sus obras se vean desprotegidas de los derechos de propiedad intelectual, debido a la facilidad que existe en nuestros días de publicar en internet una obra ajena sin la debida autorización de su autor ni del editor, provocando grandes pérdidas económicas y una ralentización en el desarrollo científico y artístico de nuestra sociedad.

Efectivamente, estamos hablando sobre delitos contra el patrimonio y contra el orden socio-económico, más concretamente sobre los delitos relativos a la propiedad intelectual, que están recogidos en nuestro Código Penal (LO 10/1995, de 23 de noviembre), en sus artículos 270 al 272.

Este presupuesto penal es lo que comúnmente se conoce como plagio. Según el Diccionario de la Lengua Española, plagiar significa "copiar en lo sustancial obras ajenas, dándolas como propias" (DLE, 2020). Este concepto, además, está legalmente definido en el artículo 270.1 CP de la siguiente manera:

> "Será castigado con la pena de prisión de seis meses a cuatro años y multa de doce a veinticuatro meses el que, con ánimo de obtener un beneficio económico directo o indirecto y en perjuicio de tercero, reproduzca, plagie, distribuya, comunique públicamente o de cualquier otro modo explote económicamente, en todo o en parte, una obra o prestación literaria, artística o científica, o su transformación, interpretación o ejecución artística fijada en cualquier tipo de soporte o comunicada a través de cualquier medio, sin la autorización de los titulares de los correspondientes derechos de propiedad intelectual o de sus cesionarios".

Como podemos observar, estamos hablando de un delito grave, puesto que cabe la posibilidad de reclusión mayor a seis meses, pero para que haya penalidad es necesario que exista ánimo de obtener un beneficio económico y que sea en perjuicio a un tercero. En otras palabras, es necesario que exista la intencionalidad de cometer el plagio.

No obstante, desde la perspectiva académica puede cometerse plagio, aun cuando no sea castigable conforme a la ley penal, ya que las obras que se exigen en el ámbito científico generalmente requieren que sean originales. Por lo tanto, aunque

el autor original haya dado permiso al autor del plagio o incluso cuando este último no obtuviere ningún tipo de beneficio económico, aun con el supuesto permiso del autor original, la autoridad competente, ya sea el profesor o el comité académico, podrá castigar igualmente al plagiador, de acuerdo con la normativa perteneciente a la institución académica o incluso mediante la costumbre, ya que ambas son fuentes de derecho tan válidas como la propia ley penal.

Debido a la seriedad de este asunto, es necesaria una aproximación sobre el plagio desde una perspectiva académica que analice sus características lingüísticas, así como los métodos presentes en la actualidad para poder detectarlo y, por último, describirlas dificultades que puede detectar el lingüista forense, imprescindible durante el proceso de detección y análisis de plagio.

Con el fin de descubrir dichas dificultades, hemos llevado a cabo un caso real, al analizar el presunto plagio de la polémica tesis doctoral de Pedro Sánchez Castejón, utilizando para ello el programa de detección de plagio más usado en la actualidad por las instituciones académicas de todo el mundo, el denominado Turnitin. Además, se ha intentado calcular su precisión en la detección de plagio y el porcentaje de coincidencia detectado que pertenece, según esta herramienta, a la bibliografía de la tesis.

2. EL PLAGIO.

La investigación sobre el plagio debe incluirse dentro de la Lingüística Forense (de aquí en adelante LF), que es un campo multidisciplinar que se incluye dentro de la Lingüística Aplicada. Es comúnmente conocida por su estrecha relación con el ámbito de la legislación, pero no podría vincularse solamente al lenguaje de la ley, sino que también guarda relación con otros campos de estudio de la lingüística, como son la sociolingüística, la fonética, la morfología, la sintaxis, la lexicología y el análisis discursivo, textual y pragmático.

También es una ciencia interdisciplinar, lo que quiere decir que trabaja en colaboración con multitud de ramas de conocimiento, tales como "la antropología, la comunicación, la criminología, el inglés, las humanidades, el derecho, la lingüística, las lenguas modernas, la filosofía, las ciencias sociales, la sociología y los estudios de traducción" (Coulthard y Johnson, 2010: 2).

Tal vez sea por su multi e interdisciplinariedad por lo que la LF tiene gran multitud de aplicaciones, de las cuales nuestra investigación se centrará en la utiliza a la lengua como evidencia, más concretamente en la investigación sobre el plagio, la

cual, aunque está en constante desarrollo, el desarrollo de nuevas técnicas computacionales y programas de ordenador ha permitido profundizar en el caso.

La metodología que suele usarse en la investigación sobre el plagio podría dividirse en dos grandes grupos, cuya denominación puede variar del autor, pero que guardan bastante semejanza. Esta dualidad metodológica, por ejemplo, Wright y Johnson la diferencia en investigación cualitativa y cuantitativa (2014:39), mientras que Nini y Grant sugiere que esta división debe ser entre estilística y cognitiva (2013:173).Algunos autores, según Wright y Johnson, abogan incluso por un método mixto: Solan y Tiersman (2004); Cotterill (2010); Turell (2010); Kredens y Coulthard (2012); Queralt y Turell (2012); Solan (2013); y Nini y Grant (2013) (Wright y Johnson, 2014: 39).

A pesar de que puede parecer que dichos programas de ordenador solo son aplicables a la investigación cuantitativa (o cognitiva), lo cierto es que también pueden servir como una herramienta muy útil, que facilita la tarea del lingüista forense cuando afronta la investigación desde una perspectiva cualitativa o estilística, como es nuestro caso.

Los estudios sobre el plagio, además, podrían relacionarse con los estudios de atribución de la autoría, ya que ambas ramas de la lingüística forense buscan peculiaridades lingüísticas con las que reconocer o identificar a un hablante determinado. "La atribución de la autoría forense concierne a la identificación de autores de documentos disputados o anónimos, que constituyen una evidencia potencial en casos legales, a partir del análisis de pistas lingüísticas olvidadas por el autor" (Johnson y Wright, 2014: 37).

Estas pistas que deja el autor podrían considerarse una huella lingüística, semejante al ADN, que es capaz de discriminar un hablante del resto por su manera de expresarse. Esta especie de huella lingüística es comúnmente conocida como 'idiolecto', y forma parte de una teoría que sigue sin resolverse.

2.1. Principales características del plagio.

Desde una perspectiva académica, no existe un consenso entre la comunidad científica para definir las distintas características y formas del plagio. No obstante, una de las taxonomías más completas que hemos podido encontrar es la ofrecida por Alzahrani, Salim y Abraham (2012: 134), que distingue el plagio literal del inteligente, y dentro de estos dos grandes grupos, encontramos distintas técnicas o subcategorías:

Por un lado, en el plagio literal, estarían recogidos:

a) la copia exacta (incluyendo documentos enteros o partes de ellos),
b) la copia aproximada (también entendidos por otros autores como parcial, e incluye la adición, la supresión, la sustitución, y la inclusión o partición de oraciones),
c) y la copia modificada (en la que se reestructura la sintaxis de la copia original o se reordena).

(Alzahrani, Salim y Abraham, 2012: 134).

Este bloque trata de distinguir la distancia de la obra plagiada con respecto a la literalidad de la copia original. Mientras que la copia exacta refleja poca intencionalidad de ocultamiento, en la copia modificada sí que hay una mayor maldad para no ser descubierto. La copia aproximada trataría también de dificultar la detección de plagio, aunque en lugar de reestructurar toda la oración, dejaría intactos ciertos elementos gramaticales, modificando principalmente los elementos semánticos del texto.

Por otro lado, en el plagio inteligente se encuentran los siguientes subgrupos:

a) la manipulación textual, incluyendo:

i. paráfrasis (léxicas o morfosintácticas)

ii. o resúmenes (del tipo de reducción oracional, de partición o inclusión de oraciones, de restructuración, de paráfrasis, y de generalización o especificación de conceptos),

b) la traducción, ya sea automática (con traducción exacta o con corpus) o manual.

c) y la adopción de ideas, donde se incluyen:

i. el significado basado en la semántica (con paráfrasis, resúmenes o traducción),

ii. la importancia basada en las secciones de un texto (contribuciones, hallazgos, resultados y discusiones),

iii. la adopción basada en el contexto (estructural o de secuencias de ideas).

(Alzahrani, Salim y Abraham, 2012: 134).

Podemos entender que el plagio inteligente constituye una aproximación a los procesos cognitivos que tratan de ocultar una copia aproximada o modificada, haciendo referencia al bloque de plagio literal.

Otras técnicas de plagio que se deben tener en cuenta a la hora de evaluar el plagio de un texto, y que no están incluidos en esta taxonomía, pueden ser:

a) La sinonimia (uso de sinónimos para tratar de ocultar el plagio);
b) La acronimia;
c) La inversión de la voz oracional (de activa a pasiva, o viceversa).

Estos nuevos ejemplos de plagio concuerdan perfectamente con la taxonomía de Alzahrani, Salim y Abraham (2012: 134) y podrían incluirse dentro de la categoría de manipulación textual de plagio inteligente, siendo la sustitución por sinónimos o la acronimia técnicas similares a la paráfrasis léxica, mientras que la inversión de la voz oracional estaría más próxima a una paráfrasis morfosintáctica, aunque también podría entenderse como copia modificada de un plagio literal, al constituir una reestructuración de la sintaxis.

Distinta a esta taxonomía, podemos encontrar las características del plagio según Guillén Nieto (2018) donde las categorías no son estrictamente lingüísticas, sino que parecen ofrecer una gradación de perjuicio ocasionado por el plagio, que pueden servir para su evaluación final. Para ello, se distinguen 7 criterios fundamentales:

1) Préstamo directo vs préstamo indirecto (copia modificada) – más malicia en la copia modificada.
2) Préstamo parcial vs préstamo total de la obra fuente.
3) Préstamo intencionado vs préstamo no intencionado.
4) Préstamo evidente vs préstamo oculto.
5) Préstamo autorizado vs préstamo no autorizado.
6) No se advierte afán de lucro vs hay afán de lucro.
7) Las obras son diferentes y claramente diferenciables vs las obras no son diferentes ni claramente diferenciables.

El préstamo directo podría entenderse como la copia literal, que, según la taxonomía anterior, podría ser exacta, aproximada y modificada (2012: 134), mientras que la clasificación de Guillén Nieto (2018) la divide en total o parcial.

Por otro lado, el préstamo indirecto (2018) se puede relacionar con el plagio inteligente, dividido en manipulación textual, traducción y adopción de ideas (2012:134).

Estos dos tipos de categorizaciones del plagio nos permite, por un lado, aplicar las distintas categorías de Alzahrani, Salim y Abraham (2012) con fines estrictamente

lingüísticos, como podría ser una investigación más profunda y exhaustiva del plagio que estamos analizando; mientras que, por otro lado, los criterios de Guillén Nieto (2018) nos sirven para evaluar lo perjudicial o malintencionado que ha podido ser el plagio para explicar nuestro análisis ante un tribunal.

2.2. PROGRAMAS DE DETECCIÓN DE PLAGIO.

Los programas de detección de plagio, también llamados programas antiplagio, son una herramienta muy útil que facilitan la labor del lingüista forense, ya que permiten un análisis del objeto de estudio más profundo y rápido. No obstante, queda todavía bastante investigación para perfeccionar su procesamiento de análisis.

Sus métodos de detección de plagio principalmente suelen dividirse en método extrínseco, que es el más común y en el que se basa prácticamente la totalidad de los programas antiplagio más conocidos, y el método intrínseco, el cual en la actualidad es solo una teoría, ya que, al menos durante esta investigación, no se ha conseguido encontrar ningún programa que utilice este método en particular.

En teoría, el método intrínseco dice que podríamos apoyarnos en patrones lingüísticos para determinar si determinados fragmentos del documento analizado podrían haber sido plagiados, ya que difieren de los rasgos estilísticos propios del autor (Zechner et al., 2009: 50)

Su hipótesis podría basarse en la idea de Nini y Grant, que dice que lo que hace a un individuo único y, por lo tanto, identificable, es su estilo, entendido como la variable elemental del comportamiento humano que incluye cada uno de los aspectos de la vida humana que requiere una elección, el cual podría encontrarse en el lenguaje mediante un análisis de selección de marcadores estilísticos que el individuo normalmente elige de su repertorio, que se genera por la variabilidad inherente de las diferencias histórico-sociológicas que distinguen la naturaleza de los seres humanos. Estos marcadores estilísticos pueden incluir diferencias estáticas, como el género o la etnia, es decir, aquellas diferencias que permanecen inalterables a través del tiempo, y diferencias dinámicas, como la edad, el trabajo, la geografía (dialecto) y la influencia de la segunda lengua, que sí que pueden variar a lo largo de la vida (Nini y Grant, 2013:174-175).

En contraposición, el método extrínseco trata de hallar coincidencias textuales entre el documento analizado y una base de datos pública o privada, que incluye resultados de búsquedas por internet, documentos especialmente protegidos para su lectura (como documentos de pago o meramente almacenados sin divulgación

pública, como un trabajo final de grado sin permiso de publicación), bibliotecas virtuales o, incluso, podrían compararse dos o más textos seleccionados manualmente.

Para mejorar la precisión de los resultados, en un futuro podrían emplearse ambos métodos conjuntamente, ya que usar en primer lugar un método intrínseco reduciría el tiempo de análisis al delimitar los fragmentos cuya autoría sería discutida, mientras que el método extrínseco trataría de determinar la autoría original de dichos fragmentos presuntamente plagiados.

Sea cual sea el método utilizado, el análisis debe ser supervisado a posteriori por un controlador humano (Zechner et al., 2009: 48) que determine si realmente ha existido plagio o no, y hasta qué punto ha podido ser intencionado. Esto se debe a los distintos tipos de plagio mencionados por Guillén Nieto (2018), descritos anteriormente en este trabajo, como, por ejemplo, cuando el autor de la obra presuntamente plagiada ha olvidado poner la referencia, aunque visualmente ha evidenciado que el fragmento constituye una cita, al entrecomillarla; o si ha citado correctamente dicha fuente en otras páginas o, por el contrario, si el fragmento discutido se ha intentado citar, aunque incorrectamente, atendiendo a la normativa de citado.

Debido a que, como hemos dicho, no se ha encontrado ningún programa que utilice el método intrínseco, hemos optado por usar un programa de método extrínseco. Los más comunes son: Antiplagiarist, Antiplagiarism, Dustball, Plagium, Plagscan, y Programme, Turnitin, Viper y WCopyfind. El problema de estos programas antiplagio es que la mayoría de ellos o bien son de pago o bien, aun siendo gratuitos, están limitados a un cierto número de palabras o de páginas, que suele ser bastante reducido.

Es por esto por lo que en este trabajo hemos decidido utilizar únicamente una herramienta, Turnitin, que no tiene límite de palabras, con el fin de determinar su eficacia para la detección de plagio, al analizar una tesis doctoral. Además, Turnitin posiblemente sea el programa antiplagio de mayor prestigio, al operar en más 15.000 instituciones, provenientes de unos 140 países, pero lo que más lo caracteriza es su gran base de datos, puesto que en 2014 ya contaba con más de 500 millones de trabajos indexados (Turnitin, 2020).

3. OBJETIVOS, PREGUNTAS Y METODOLOGÍA DE LA INVESTIGACIÓN.

3.1. Objetivos y preguntas de investigación.

Este estudio pretende analizar la tesis del actual presidente de gobierno de España, Pedro Sánchez Castejón, para determinar si ha existido plagio, utilizando la herramienta Turnitin como apoyo para detectar los fragmentos discutidos, pero finalmente calculando el porcentaje exacto de presunto plagio mediante regla de tres.

El hecho de elegir esta obra se debe a la polémica noticia de 2018, a raíz de la publicación de Javier Chicote en el periódico *ABC* que decía con gran rotundidad que, tras analizar el documento, utilizando herramientas de detección de plagio, Pedro Sánchez había cometido varios plagios en su tesis, al haber copiado párrafos enteros de otros autores, entre los que menciona a los economistas Julio Cerviño y Jaime Rivera (Chicote, 2018).

Tras esta noticia, varios periódicos se hicieron eco de la misma y el actual presidente se defendió públicamente, alegando que las fuentes sí que fueron citadas. Sin embargo, este hecho pasó a ser un arma política que varios diputados utilizaron para atacar a la figura de Sánchez, a pesar de que ningún periódico había descrito el método científico que se había usado para justificar dicha acusación, sino que se limitaban a dar porcentajes de presunto plagio obtenidos con herramientas, tales como Turnitin.

Por otro lado, también trataremos de medir la precisión de este programa antiplagio a partir de una muestra de 10 referencias, que tomaremos aleatoriamente del listado de Turnitin, cuyos fragmentos serán clasificados de acuerdo con una categorización propia que describiremos más adelante.

Tras la realización de este trabajo, se tratará de responder, desde una perspectiva académica, las siguientes preguntas de investigación:

a) ¿Ha existido plagio académico en la tesis de Pedro Sánchez? Si es así, ¿cuál es el grado de malicia existente en esta conducta?

b) ¿Se puede determinar la existencia de plagio en una tesis doctoral a través de la herramienta Turnitin? ¿Qué limitaciones existe en una investigación de este tipo?

3.2. Metodología y procedimiento.

Es necesario, en primer lugar, definir los conceptos de referencia y fragmentos indicados por Turnitin, ya que pueden ser confusos. Por un lado, una referencia es la fuente de información de la que supuestamente el autor de la obra analizada ha

podido copiar parte o la totalidad de su contenido. Turnitin las recoge en un listado que aparece a la derecha del programa y las ordena por el porcentaje de coincidencia con la obra analizada. Por otro lado, un fragmento corresponde a cada una de las partes de una referencia que coinciden con la obra analizada.

En primer lugar, se llevará a cabo un análisis automático de la tesis de Pedro Sánchez mediante el programa de detección de plagio, Turnitin, con el fin de obtener suficientes datos para realizar los análisis subsiguientes.

Luego, tomaremos una muestra de 10 referencias aleatorias detectadas por Turnitin como coincidentes, y clasificaremos cada uno de los fragmentos de acuerdo con nuestra categorización, que consta de 4 grupos:

1) referencias bibliográficas: cualquier fragmento que se refiera al título completo de una referencia bibliográfica que sirve como fuente para el texto analizado.

2) expresiones fijas: incluye grupos de series de palabras que no pueden ser considerados como plagio por su conocimiento general ("Presidente del Gobierno") o su función estrictamente discursiva ("con la entrada en vigor de la Ley"), ni como una referencia bibliográfica, porque solo forma parte de otra referencia bibliográfica ("Directiva del Parlamento Europeo y del Consejo por el que se modificaba la Directiva 2003/87/CE", donde "la Directiva 2003/87/CE" solo forma parte de otra referencia).

3) con problemas de fecha o de acceso al documento: ya sea fragmentos que pertenecen a una referencia publicada después de la tesis de Pedro Sánchez o cuya fecha no puede encontrarse; o bien cuya referencia no se encuentra disponible para su visualización o ya no se puede acceder de ningún modo, porque el contenido en internet se ha eliminado, aunque permanece en la base de datos de Turnitin.

4) constituyentes de un posible plagio: fragmentos que contienen un grado de similitud sintáctica y léxica alta, que exponen un tema similar y su citado no esté claro o sea incorrecto.

En aquellos casos en que un mismo fragmento pueda incluirse en dos grupos, para este análisis siempre prevalecerá el tercer grupo (con problemas de fecha o de acceso al documento), puesto que no serviría de nada incluirlas en otra categoría si no podemos comprobar que la fuente fue escrita antes que la tesis analizada.

Una vez categorizados todos los fragmentos, se calculará el porcentaje de la frecuencia de cada categoría de la muestra analizada, con el fin de determinar la

precisión de detección de plagio que tiene Turnitin, así como también averiguar qué tipo de fragmentos detecta esta herramienta como coincidentes, aunque no constituyen un presunto plagio, para observar mejor las verdaderas limitaciones de Turnitin.

El hecho de tomar solo 10 referencias para la muestra se debe a que sería relativamente imposible para este trabajo categorizar la totalidad de los datos, puesto que existen cientos de posibles referencias detectadas por la herramienta, las cuales incluyen, por lo general, entre 20 y 60 fragmentos cada una, aunque hay casos que van de 10 a 200 de estos, y también porque, aparentemente, existe la tendencia de que los datos se repiten.

Luego, analizaremos el resto de las referencias indicadas por Turnitin, distintas a la muestra analizada previamente, para conservar únicamente aquellos fragmentos que podrían considerarse como presunto plagio, excluyendo del listado todas las demás referencias que solo contienen expresiones fijas o cuya fecha de publicación sea posterior a la de la tesis o de acceso restringido. No obstante, si el acceso fuera restringido, pero existieran sospechas de un posible plagio, se llevaría a cabo una búsqueda externa a Turnitin para encontrar la fuente, y así poder analizar posteriormente tales fragmentos.

Finalmente, se analizarán todos los posibles casos de plagio detectados para determinar si están debida o incorrectamente citados y, si están incorrectamente citados, se analizarán cualitativamente y se calculará el porcentaje exacto de presunto plagio correspondiente a la tesis.

4. RESULTADOS.

4.1. Resultados del análisis de la muestra.

En lo relativo al primer análisis realizado sobre 10 referencias aleatoriamente escogidas del listado de Turnitin, los resultados indican que, de un total de 209 fragmentos contenidos en esas 10 referencias, 31 de ellos son relativos a referencias bibliográficas, lo que se traduce en un 14,83% de los casos.

Cabe recordar que para que un fragmento pueda pertenecer a esta categoría debe constituir una referencia bibliográfica completa, que puede ser la referencia a un artículo académico, el nombre de un texto normativo, el nombre de un plan oficial del gobierno o cualquier otro tipo de referencia que aparezca completa y que pueda ser considerada una unidad completa independiente, que no parte de otra referencia.

Si por el contrario la referencia bibliográfica es parcial (pertenece a otra referencia bibliográfica), ésta se incluiría en la categoría de expresiones fijas:

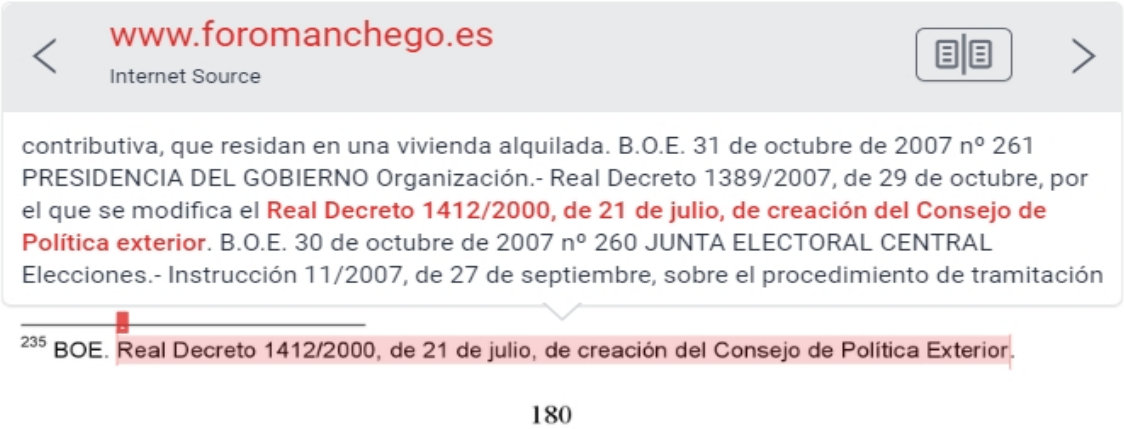

Imagen 1: Ejemplo de expresión fija. Fuente: elaboración propia, a partir de los datos obtenidos de Turnitin.

En otros casos, se observa también cómo en ocasiones se repiten determinados fragmentos en el documento analizado, mientras que en la fuente original aparece una sola vez, lo que provoca que el porcentaje de presunto plagio aumente erróneamente.

Un ejemplo de esto puede ser el que se observa en la siguiente imagen, donde "The expanding role of cities in international politics" aparece una sola vez en la fuente detectada por Turnitin, y nueve en la tesis, aunque en la imagen solo se observan dos ocurrencias:

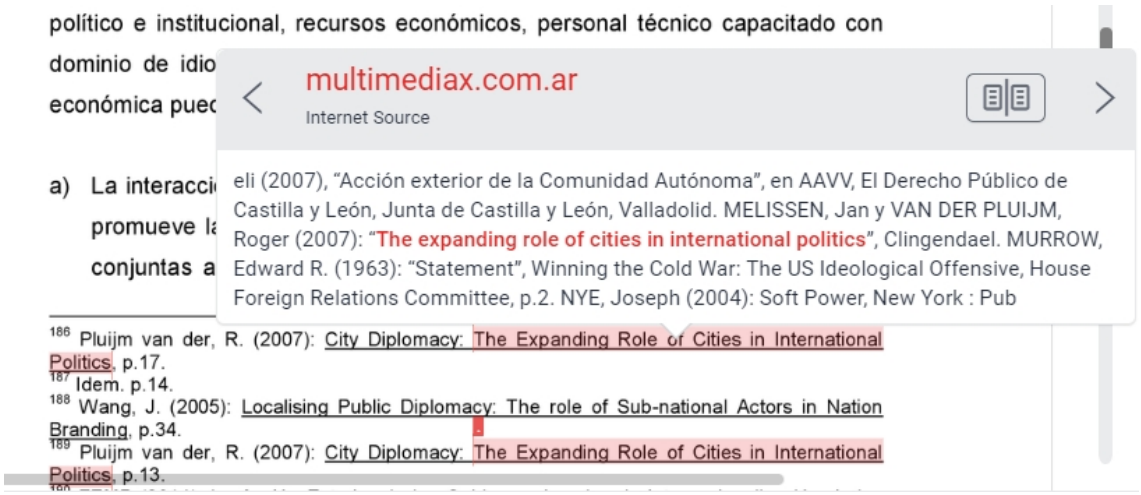

Imagen 2: Ejemplo de la categoría "referencia bibliográfica", donde un mismo fragmento de la referencia aparece más de una vez en la tesis analizada. Fuente: Elaboración propia, a partir de los datos obtenidos de Turnitin.

La segunda categoría analizada se refiere a las expresiones fijas resaltadas por Turnitin como posible plagio. El número de fragmentos que pudieron categorizarse aquí es de 24, lo que corresponde al 11,48% del total de la muestra.

Esta categoría no llega a constituir un posible plagio puesto que el número de palabras coincidentes entre la fuente indicada y el texto analizado suele ser de 3 a 5, y considerar como plagio a un número tan reducido de palabras debería considerarse un error, puesto que son simples sintagmas nominales o colocaciones, que muchas veces no guardan ni siquiera relación temática con la tesis. Además, una simple búsqueda en Google de, por ejemplo, "Presidente del Gobierno" nos da casi 36 millones de resultados.

El tercer grupo de nuestra categorización es el más frecuente y tiene que ver con aquellas referencias a las cuales no se puede acceder o cuya fecha es posterior a la tesis. Este grupo abarca 154 casos, y corresponde con el 73,68% del total de los fragmentos. Esto quiere decir que casi 3 de cada 4 fragmentos, que Turnitin indica como posible plagio, o bien no pueden comprobarse o bien fueron publicados después que la tesis, por lo que tienen que ser descartados.

Por último, tenemos el grupo de posible plagio, cuyos resultados indican que, de las 10 referencias de la muestra seleccionada, hemos encontrado 6 fragmentos discutidos cuyo acceso era imposible desde esta herramienta, por lo que, para el análisis de la muestra, tuvieron que ser categorizados en el grupo (3). Si igualmente estos 6 fragmentos los hubiéramos categorizado en el grupo de posible plagio (4), el porcentaje habría sido del 0,95% sobre el total de la muestra.

No obstante, llevamos a cabo un análisis externo a Turnitin de tales fragmentos para determinar si ha habido plagio en ellos, como veremos a continuación.

4.2. Análisis de posibles casos detectados en la muestra y en el resto de la tesis.

4.2.1. Primer caso.

Al analizar los 6 fragmentos discutidos del análisis de la muestra, descubrimos que 4 de ellos sí estaban correctamente citados, mientras que los otros 2 podrían considerarse como presunto plagio. Debido a que su acceso desde Turnitin era imposible, decidimos realizar una búsqueda externa a esta herramienta a través de Google y encontramos que estos dos fragmentos discutidos podrían pertenecer al artículo de Julio Cerviño y Jaime Rivera Camino (2007), titulado "La Globalización de las marcas españolas: liderazgo y notoriedad internacional".

Cabe resaltar que es cierto que en la página 286 hay una referencia bibliográfica a pie de página que nombra a Cerviño, pero esta referencia corresponde con un artículo completamente distinto, ya que fue publicado junto a Cubillo en 2005.

Por lo tanto, Pedro Sánchez no ha citado correctamente el párrafo que abarca las páginas 285 y 286, cuya autoría es discutida, porque cabe duda de que dicha fuente no se ha citado correctamente.

Como conclusión, este fragmento podría considerarse un presunto plagio, ya que la fuente no ha sido correctamente citada, y el grado de similitud léxica y sintáctica es alto, comenzando con un posible plagio literal, que dice así: "Una de las principales fuentes de creación Marca País son las propias marcas comerciales y corporativas" (Sánchez 2012: 285) y (Cerviño y Rivera 2007: 128), seguido de una posible copia parcialmente reflejada en la tesis, como veremos a continuación.

Mientras que en el artículo Cerviño y Rivera (2007:128) encontramos:

> "La globalización del comercio hace que las empresas y productos, con sus marcas globales o locales, circulen internacionalmente y se encuentren en muchos países. La inmensa mayoría de los consumidores compran productos extranjeros constantemente, de países donde nunca han estado, pero que conocen y perciben por imágenes, comentarios, noticias, relatos, etcétera. A partir de su experiencia con estas marcas y empresas, los consumidores conforman todo un mundo de significados relacionados con su país de origen".

La tesis de Pedro Sánchez (2012: 285-286) contiene lo siguiente:

> "En mercados globales, los consumidores y usuarios compran bienes y servicios fabricados en países de los cuales sólo conocen por imágenes, comentarios, noticias, etc. A partir de sus experiencias con esas empresas y marcas, los consumidores conforman todo un mundo de significados relacionados con su país de origen".

Como podemos observar, la presunta copia con modificación de la tesis incluiría, además de ciertos fragmentos idénticos, las siguientes técnicas:

a) paráfrasis ("La globalización del comercio", reemplazado en la tesis por "En mercados globales"; o "La inmensa mayoría de los consumidores compran productos extranjeros constantemente, de países donde nunca han estado, pero que conocen y perciben por imágenes, comentarios, noticias, relatos, etcétera", sustituido en la tesis por "los consumidores y usuarios compran bienes y servicios fabricados en países de los cuales sólo conocen por imágenes, comentarios, noticias, etc.");
b) supresión de la posible fuente original ("hace que las empresas y productos, con sus marcas globales o locales, circulen internacionalmente y se encuentren en muchos países", "donde nunca han estado", "y perciben", o "relatos");

c) alternancia del número ("su experiencia" por "sus experiencias");

d) y reordenación ("estas marcas y empresas" por "esas empresas y marcas").

Sin embargo, este presunto plagio podría no ser intencional, ya que sí que cita esta fuente en otras páginas (293 y 403), por lo que sería evidente que ha usado dicha fuente, que no ha tratado de ocultarlo completamente, a pesar de haber usado ciertas técnicas, como las descritas anteriormente, y que ambas obras son, en su totalidad claramente distintas y diferenciables.

Finalmente, podemos concluir que este posible plagio abarca una extensión de 64 palabras, que hacen un total del 0,07% del total de la tesis (90.280 palabras).

4.2.2. Segundo caso.

Otro ejemplo de posible plagio encontrado al analizar el resto de la tesis puede ser la gran similitud que se da en las páginas 295 y 296 de la tesis, comparado con Iglesias Onofrio y Molina Rabadán (2008: 116). En ambos documentos se habla sobre las recomendaciones del Proyecto Marca España (2003).

Aunque es cierto que Pedro Sánchez incluye una recomendación adicional, que corresponde con el apartado de (j) de su clasificación, la cual no se encuentra recogida en Iglesias Onofrio y Molina Rabadán (2008), el resto de las recomendaciones aportadas en la tesis guardan mayor semejanza con el artículo de 2008 que con el mismo Proyecto Marca España, por lo que se deduce que el artículo es la posible fuente original del fragmento presuntamente plagiado de la tesis.

Cabe señalar que es cierto que Pedro Sánchez menciona tanto al Proyecto Marca España como al artículo de Iglesias Onofrio y Molina Rabadán en un pie de página, que encontramos en la página 295, pero no los cita correctamente, entendiéndose que los fragmentos han sido escritos por Pedro Sánchez, siendo estas referencias solo una fuente para aumentar la información aportada, como se observa a continuación:

[406] Para un análisis más detallada, véase: Informe: Proyecto Marca España. Asociación de Directivos de Comunicación, Instituto Elcano de Estudios Internacionales y Estratégicos. Foro de Marcas Renombradas Españolas. ICEX. pp. 23-32. Mayo, 2003; y Iglesias Onofrio, M y Molina Rabadán, D. (2008): La Estrategia Marca País en la sociedad informacional, pps.109-26.

295

Imagen 3: Pie de página de la tesis sobre el Proyecto Marca España, en la página 295.

Así pues, en las siguientes imágenes vamos a observar los fragmentos discutidos, con el fin de analizar sus diferencias y comprobar si ha habido plagio:

En el Informe se recomendó[15]:

- Constituir un observatorio permanente de la imagen exterior -Observatorio Permanente de la Imagen de España en el Exterior (OPIEX)- con la finalidad de hacer un seguimiento continuo de las percepciones y actitudes ante nuestro país en todo el mundo a través de seminarios, estudios e investigaciones de diverso tipo.
- Establecer un servicio de alerta y comunicación para detectar y atajar situaciones de crisis de la imagen en algún país y elaborar unas líneas estratégicas de posicionamiento y comunicación de la Marca-España
- Potenciar el aspecto comercial, y por lo tanto, de la denominación Marca-España.
- Coordinar las acciones de las instituciones públicas y privadas que hoy día crean imagen de España.
- Dar un mayor peso en el servicio exterior a la formación en temas de diplomacia pública.
- Poner en marcha un programa de relaciones públicas y de *lobbying* en el exterior al servicio de los intereses de España. Son ejemplos, las dos giras del presidente Aznar por los estados de USA con una mayor presencia de población hispana en julio y en septiembre de 2003, y el lanzamiento de la campaña *Spain: the friend in Europe* con inserciones en los principales medios norteamericanos[16]. También la asistencia de los Reyes a las Cumbres Iberoamericanas y eventos en países latinoamericanos;
- Reforzar y coordinar la acción cultural exterior fortaleciendo el accionar del Instituto Cervantes con alianzas estratégicas y financiando viajes y estadías de estudio en España a extranjeros, como lo realiza la Fundación Carolina o el Programa Erasmus de la UE.
- Promocionar las buenas prácticas de imagen y la figura de los embajadores de la marca España oficiales y "oficiosos" (deportistas, artistas, profesionales) mediante convocatorias de premios o incentivos a aquellos que hayan contribuido a la mejora de la imagen de España.
- Aprovechar los eventos culturales, deportivos y artísticos de nivel internacional que tengan lugar en el país como la Exposición Universal de Sevilla y las Olimpiadas de Barcelona de 1992, la candidatura de Madrid como sede de los Juegos Olímpicos de 2012, etc. para reforzar y difundir la nueva imagen del país[17].

Imagen 4: Fragmento de Iglesias Onofrio y Molina Rabadán (2008: 116), que presuntamente sirvió como fuente para la tesis.

(a) Crear un Observatorio Permanente de la Imagen de España en el Exterior (OPIEX) para hacer un seguimiento continuado de las percepciones y actitudes ante nuestro país del mundo a través de estudios periódicos y la celebración de jornadas, seminarios, etc.

(b) Poner en marcha un servicio de alerta para detectar y atajar situaciones de crisis de la imagen española en algún país e implementar líneas estratégicas de posicionamiento y comunicación de la Marca España.

(c) Potenciar el aspecto comercial de la Marca España.

(d) Coordinar las acciones de instituciones públicas y privadas que hoy día crean imagen de marca de España.

(e) Dar un mayor peso al servicio exterior en la formación en temas de diplomacia pública.

(f) Implementar programas de Relaciones Públicas y lobby en el exterior a servicio de los intereses de España.

(g) Reforzar y coordinar la acción cultural exterior con el Instituto Cervantes como principal referente y la Fundación Carolina, financiando, por ejemplo, viajes a extranjeros para aprender español.

(h) Promocionar las buenas prácticas de imagen y la figura de los embajadores de la Marca España oficiales y "oficiosos" (deportistas, artistas, profesionales, etc.) mediante convocatorias de premios o incentivos a aquellos que hayan contribuido a la mejora de la imagen de España.

(i) Aprovechar los eventos culturales, deportivos, artísticos de nivel internacional para reforzar y difundir la imagen del país.

Imagen 5: Fragmento de la tesis (páginas 295 y 296) donde se encuentra resaltado el texto coincidente con Iglesias Onofrio y Molina Rabadán (2008: 116).

Al comparar ambos textos, se puede observar que Pedro Sánchez tan solo ha intercambiado algunas palabras gramaticales por sinónimos (como "para", sustituyendo "con la finalidad de"; o "e implementar", que sustituye "y establecer"), o ha resumido algunos párrafos mediante supresión de contenido (como se observa en los párrafos (c) e (i) en la tesis).

Otra estrategia que ha podido usar para disimular el presunto plagio podría ser el hecho de incluir una enumeración con letras, en lugar de usar la viñeta que usa Iglesias Onofrio y Molina Rabadán (2008: 116).

Como conclusión, Pedro Sánchez ha podido ocultar que ha tomado prestado estos fragmentos del artículo de 2008 por dos motivos: porque ha podido emplear técnicas de plagio literal parcial con modificación textual y porque no ha citado correctamente estas recomendaciones, una conducta que no se entiende, puesto que sí que ha incluido una referencia a dicho artículo.

No obstante, hay que ser precavidos al calificar este fragmento como presunto plagio, ya que ha podido no ser intencional, puesto que en otras páginas sí que cita correctamente a estos autores, como se observa en las páginas 283 y 290 de la tesis, por lo que podría ser evidente.

Además, si contamos las palabras de estos apartados supuestamente plagiados, nos da un total de 225, lo que constituye el 0,249% del total de la tesis.

4.2.3. Tercer caso.

Este último caso ha sido encontrado al analizar el resto de las referencias del listado de Turnitin y guarda cierta semejanza con el caso anterior, aunque éste tiene una dificultad añadida, ya que la posible fuente original está escrita en inglés (y no en español, como ocurre en el Proyecto Marca España).

En este caso, la presunta fuente original corresponde con el artículo de Saner y Yiu (2001: 12-16), que ha sido parafraseado en García Rebollar (2010: 24), el cual ha traducido el artículo al español en su obra.

Aquí, Pedro Sánchez menciona a los autores originales (Saner y Yiu, 2001), a quienes parece citar, pero incluye un pie de página que dirige a García Rebollar (2010). Esto crea cierta confusión, porque a quien cita realmente es a Rebollar (2010).Para que se entienda mejor, éste es el pie de página al que nos referimos:

Saner y Yiu (2001)[53] diferencian la diplomacia económica del resto de acciones diplomáticas en función de los actores implicados en la misma (ver cuadro 1). En este sentido, para los autores, son los Estados o agentes económicos privados –empresas, ONG- y la distinta naturaleza de los intereses que promueve cada uno de ellos en su labor diplomática (el interés general del país en cuestión o de una determinada empresa) lo que distingue a la diplomacia

[49] Bassols, R. (1995): La diplomacia al servicio de la economía, pps: 195-197.
[50] Azuelos, M. (2011): Les évolutions de la diplomatie économique américaine, p. 12.
[51] Chavagneux, C. (1999): La diplomatie économique: plus seulement une affaire d'États, p.33.
[52] Morillas, J. (2000): La diplomacia económica, pps.41-44.
[53] Utilizaremos las definiciones de cada uno de los términos traducidos en el trabajo de Rebollar G. (2010): Notas sobre diplomacia económica, pps. 25-34.

42

Imagen 6: Pie de página donde se menciona a Saner y Yiu (2001), pero la referencia redirige a García Rebollar (2010). Encontrado en la tesis de Pedro Sánchez (2012: 42).

Consecuentemente, la obra que vamos a comparar con la tesis será la de García Rebollar (2010), ya que es presuntamente la fuente que utiliza:

– Diplomacia económica: engloba el trabajo de los funcionarios de ministerios sectoriales responsables de organismos multilaterales de economía y comercio, y a los funcionarios que realizan el seguimiento de las políticas económicas y comerciales de otros países.

– Diplomacia comercial: incluye principalmente el trabajo de las oficinas comerciales en el exterior en la promoción de exportaciones y la captación de inversiones extranjeras.

– Diplomacia empresarial: es la actividad de los ejecutivos de las empresas multinacionales en sus relaciones exteriores, tanto en las tareas realizadas en la sede central como en las sucursales en el exterior.

– Diplomacia del negocio: es la parte del trabajo de los ejecutivos de una multinacional que tiene relación con los Gobiernos con el fin de influir sobre una legislación concreta, participar en concursos públicos o realizar campañas de imagen ante los Gobiernos, las ONG u otros medios de opinión.

– Diplomacia de ONG nacional: cuando su papel se extiende fuera del país.

– Diplomacia de ONG multinacional: se trata de redes bien conectadas en los principales países que llegan a ser prescriptores de opinión y a condicionar las políticas de los Gobiernos y de las empresas (por ejemplo, Greenpeace).

Imagen 7: Fragmento del artículo de García Rebollar (2010: 24) que pudo haber servido de fuente para la tesis

- La diplomacia económica: incluye la tarea desempeñada por los funcionarios de los Ministerios encargados de las relaciones con los organismos multilaterales de economía y comercio, y la de los funcionarios que realizan el seguimiento de las políticas económicas y comerciales del país.

- La diplomacia comercial: recogería la labor realizada por las Oficinas Comerciales en el exterior en la promoción de las exportaciones y la captación de inversiones extranjeras.

- La diplomacia corporativa: la actividad de los ejecutivos de las empresas multinacionales encargados de las relaciones externas tanto en la labor que realizan en la sede central como en las oficinas en el exterior.

- La diplomacia de los negocios: la acción realizada por las empresas ante las Administraciones Públicas y Organismos Internacionales con el fin de influir en legislaciones concretas, concursos públicos, etc.

- La diplomacia de la ONG multinacional y nacional: redes conectadas en distintos países que tratan de condicionar las decisiones de gobiernos, organismos multilaterales o regionales, y de empresas multinacionales.

Imagen 8: Fragmento de la tesis de Pedro Sánchez (2012: 43), cuyos fragmentos resaltados son los que coinciden con el artículo de García Rebollar (2010: 24).

Si comparamos ambas obras, observamos que Pedro Sánchez ha podido tratar nuevamente de ocultar el préstamo, ya que estos párrafos guardan un grado de similitud alto, donde tan solo distan algunas estructuras sintácticas (como la inclusión del artículo femenino singular, "la", antes de cada apartado, algo que no aparece en el artículo de 2010) y morfológicas (como el plural para "legislaciones concretas", cuando en la posible fuente original aparece en singular).

Cabe resaltar también el nombre del tercer apartado, el cual aparece en la posible fuente original como "diplomacia empresarial", mientras que Pedro Sánchez lo denomina "la diplomacia corporativa", lo que podría considerarse una paráfrasis léxica o una sustitución sinonímica.

También encontramos una serie de sinónimos, tal vez, para tratar de ocultar este préstamo. Para que sea más fácil de entender, entre paréntesis mostramos cómo aparecen en el artículo de García Rebollar (2010). Algunos de los sinónimos encontrados son: "la tarea" o "la labor" ("el trabajo"); "incluye" ("engloba"); "recogería" ("incluye"); "relaciones externas" ("relaciones exteriores"); y "las Administraciones Públicas y Organismos Internacionales" ("los Gobiernos").

Este posible préstamo literal incluye también supresiones, tales como "sectoriales" (visible en el primer apartado del artículo de Rebollar, mientras que se habría eliminado en la tesis); "principalmente" (que aparece en el apartado segundo del artículo, y no en la tesis) o de ejemplos de organizaciones internacionales, como "Greenpeace", que tampoco aparece en la tesis.

Como conclusión, se podría decir que también ha existido un posible caso de plagio en estos determinados fragmentos, puesto que, aunque sí menciona a la fuente original escrita en inglés, habiendo observado esa fuente, podría decirse que los

fragmentos de la tesis guardan más similitud con el artículo de García Rebollar (2010) que con el de Saner y Yiu (2001), por lo que no habrían sido correctamente citados.

No obstante, este posible plagio podría no haber sido intencional, porque en cierto modo ha tratado de citar la fuente, pero no ha seguido el procedimiento correcto de citado, lo que provoca cierta confusión.

Por último, hay que destacar que la suma de las palabras de estos fragmentos hace un total de 165 palabras, lo que correspondería con un 0,182% del total de la tesis.

4.3. Resultado final del análisis de presunto plagio.

Como conclusión, tras el análisis completo de la tesis, el porcentaje total que puede ser considerado como presunto plagio corresponde a un 0,501% del total de la obra (454 palabras supuestamente plagiadas de las 90.280 que contiene la tesis). Este porcentaje es la suma de los tres casos detectados como presunto plagio.

Cabe recordar que tras analizar estos fragmentos, podríamos concluir que aparentemente no ha existido mala fe, ya sea porque ha intentado citarlos, aunque incorrectamente (casos 2 y 3) o bien porque los ha citado en otras páginas, pero no en la discutida (primer caso), lo que evidencia el uso de dichas fuentes, al menos, a lo largo de la tesis, por lo que en ese aspecto no se han ocultado completamente, por lo que más que un plagio, esto podría considerarse un error humano que ha surgido a la hora de citar dichas fuentes de manera apropiada.

5. DISCUSIÓN DE LOS RESULTADOS.

Habiendo procesado los resultados de los distintos análisis realizados en este trabajo, podemos responder a las preguntas de investigación, formuladas anteriormente.

Pregunta de investigación 1: ¿Ha existido plagio académico en la tesis de Pedro Sánchez? Si es así, ¿cuál es el grado de malicia existente en esta conducta?

En primer lugar, con respecto al supuesto plagio, podemos concluir que la tesis doctoral de Pedro Sánchez no es completamente un plagio en sí misma, pero es cierto que hay 3 fragmentos que hemos analizado que pueden ser constituyentes de un posible caso de plagio, debido a que las referencias que hemos encontrado se asemejaban en un alto grado de similitud léxica y sintáctica, que estas fueron publicadas antes de la tesis y que son obras que Pedro Sánchez ha usado para el

contenido de otras páginas de su tesis, aunque citándolas correctamente. Por lo tanto, Pedro Sánchez conocía la existencia de estas obras, y las ha usado (por lo general, correctamente, pero indebidamente en estos 3 casos particulares).

No obstante, el hecho de haber citado anteriormente estas obras, puede significar que esta conducta ha podido ser involuntaria, por lo que el grado de malicia sería mínimo.

Pregunta de investigación 2: ¿Se puede determinar la existencia de plagio en una tesis doctoral a través de la herramienta Turnitin? ¿Qué limitaciones existe en una investigación de este tipo?

Se puede decir que es posible determinar la existencia de plagio en una tesis doctoral a través de la herramienta Turnitin, aunque con un coste elevado de tiempo y esfuerzo, debido a la gran magnitud de una obra de tal envergadura y complejidad. Consecuentemente, Turnitin es una herramienta de gran utilidad por llevar a cabo un proceso automático que reduce una gran cantidad de tiempo y esfuerzo al investigador. Sin embargo, el plagio solo puede detectarse mediante una intervención humana, que debe ser un humano con criterio científico y con formación en lingüística forense, con el fin de que los resultados obtenidos puedan ser más fiables y rigurosos, que son condiciones exigidas siempre ante un tribunal.

A esto, se le suman todas las limitaciones que hemos encontrado durante la realización de esta investigación, entre las que se incluyen: la ausencia de un procedimiento convencionalizado para el desarrollo de una investigación de este carácter; la dificultad de categorización de una gran masa de datos a partir de un análisis manual e individualizado; la imposibilidad de excluir fragmentos de una referencia indicada por Turnitin, lo que supone una limitación en la exactitud del porcentaje de coincidencia que muestra esta herramienta; y la necesidad de una extrema rigurosidad científica, por la seriedad del hecho y el objeto que se ha analizado.

6. CONCLUSIONES.

Tras la realización de este trabajo, se ha podido comprobar la infalible necesidad de un controlador humano, que debería ser especialista en lingüística forense, con el fin de analizar exhaustivamente cada uno de los datos provenientes de herramientas de detección de plagio y poder determinar la existencia y el porcentaje real de plagio, así como también poder hacer una valoración cualitativa y no puramente cuantitativa del supuesto plagio.

La condición de que el controlador humano debería de ser un lingüista forense se debe a que este estará más especializado en el contexto legal y académico del plagio, en cada uno de los componentes que estructuran y dan significado al lenguaje, y en que conocerá de primera mano los requisitos científicos que se le exigen en las pruebas que se aporten ante un tribunal.

También se han podido observar una serie de limitaciones que han surgido tanto de la investigación como del uso de la herramienta Turnitin:

En primer lugar, no existe ningún criterio evaluador, convencionalizado por las instituciones académicas, para la realización de esta tarea. Gran parte de este análisis, por ende, recae sobre la intuición.

En segundo lugar, desde nuestro punto de vista, Turnitin tiene multitud de limitaciones que dificultan o ralentizan en gran medida el análisis exhaustivo sobre el plagio. Entre estas limitaciones se encuentran:

a) La imposibilidad de descartar automáticamente obras bibliográficas (referencias del listado de coincidencias) que han sido publicadas posteriormente al documento analizado.
b) La imposibilidad para detectar coincidencias textuales en obras escritas en otra lengua, puesto que solo detecta coincidencias en palabras morfológicamente idénticas.
c) La imposibilidad de ver la fecha de publicación de algunas fuentes cuyo acceso está restringido, así como el acceso a los mismos.
d) La imposibilidad de descartar manualmente determinados fragmentos de una misma referencia, ya que Turnitin solo permite excluir la referencia completa, pero no sus fragmentos. Esto nos lleva a la conclusión de que el porcentaje de presunto plagio que indica Turnitin, posiblemente en la mayoría de los casos, no sea real, sino una estimación a la alza, puesto que una referencia que pueda considerarse un presunto plagio puede contener fragmentos que no lo sean, lo que hace aún más necesaria la presencia del controlador humano.
e) La duplicidad de fuentes en la base de datos, que provoca que una misma referencia bibliográfica pueda encontrarse más de una vez en el listado de referencias de Turnitin, a veces con distinto porcentaje de coincidencia, si una de las obras duplicadas se encuentra parcialmente almacenada en la base de datos.

Esto nos lleva a cuestionar la fiabilidad existente en las herramientas antiplagio y si debemos confiar plenamente en ellas o, por el contrario, dudar también de los resultados de este tipo de herramientas y cuestionarnos su verdadera utilidad, así como sus limitaciones, con el fin de optimizar su procesamiento y obtener resultados más fiables.

En tercer lugar, la última y mayor dificultad atiende a la velocidad de procesamiento humano para el análisis de las referencias detectadas por esta herramienta. El hecho de tener que revisar uno a uno cada fragmento, que Turnitin detecta como coincidente y que al analizar una tesis doctoral se cuentan por miles, es un proceso muy lento y que el mismo software es incapaz de enjuiciar por sí solo.

Consecuentemente, se debe abogar por el desarrollo de una herramienta de detección de plagio que combine ambos métodos (extrínseco e intrínseco), con el fin de mejorar la precisión de los programas de detección de plagio, que acabaría en una mayor calidad de análisis y en una reducción del tiempo requerido en el control humano. Combinar estos métodos supondrían un avance sin precedentes en el estudio sobre el plagio, puesto que esta herramienta pasaría de centrar su análisis en la mera coincidencia léxica a centrarlo entonces en la propia diferencia estilística del texto, que constituye su huella digital, para luego comparar los fragmentos discutidos con la base de datos y así obtener su posible fuente referencial. Esto ayudaría a encontrar nuevos plagios o fragmentos que no fueron escritos por el autor de la obra que las actuales herramientas no son actualmente capaces de encontrar.

Como conclusión a esta investigación, solamente mediante el análisis de esta herramienta no se puede determinar que ha existido plagio, puesto que las únicas 3 referencias coincidentes con la tesis que pueden corresponder con un presunto caso de plagio parcial (0,501% del total de la tesis) han tenido que ser analizadas por otro método, aunque Turnitin ha sido útil para detectarlas en el texto.

Desde la perspectiva académica, a pesar de que hayamos podido encontrar 3 fragmentos de la tesis que podrían formar parte de un posible caso de plagio (cuya malicia puede ser discutida, pero aparentemente no sería intencional, ya que en otras páginas sí que ha citado las mismas fuentes), no implica que la tesis doctoral de Pedro Sánchez pueda ser considerada en su totalidad un plagio, ya que, relativamente, el posible plagio que se ha encontrado puede considerarse ínfimo, en comparación con la inmensidad de un documento académico, como es una tesis doctoral.

Por último, cabe resaltar que se debe ser más precavido a la hora de calificar una obra como plagio por la seriedad que esto conlleva, ya que las implicaciones que tiene la conducta del plagiador no son solamente punibles desde la perspectiva académica, por la que se puede suspender, expulsar o incluso, tal vez, retirar el título obtenido, sino que también pueden ser penales desde el ámbito legal, pudiendo imponerse un castigo que va desde "una multa de doce a veinticuatro meses" hasta "la pena de prisión de seis meses a cuatro años", según el artículo 270.1 CP.

BIBLIOGRAFÍA.

Alzahrani, S., Salim, N., y Abraham, A. (2012). Understanding plagiarism: Linguistic patterns, textual features, and detection methods. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* 42(2), 133-149. Disponible en: http://sip04.softcomputing.net/smcc2011.pdf [Con acceso el 10 de mayo de 2020].

Asociación de Directivos de Comunicación, ICEX, DIRCOM y Real Instituto Elcano de estudios internacionales y estratégicos. (2003). *Proyecto marca España*. Madrid: Cyan, Proyectos y Producciones Editoriales, S.A. Disponible en: http://www.realinstitutoelcano.org/wps/wcm/connect/b9f4f0804f0195c7887cec3170baead1/Informe_Proyecto_Marca_Espana.pdf?MOD=AJPERES&CACHEID=b9f4f0804f0195c7887cec3170baead1 [Con acceso el 10 de mayo de 2020].

Cerviño, J., y Rivera Camino, J. (2007). La Globalización de las marcas españolas: liderazgo y notoriedad internacional. *Información Comercial Española*, 839, 121-139. Disponible en: https://e-archivo.uc3m.es/bitstream/handle/10016/7281/ICE_839_121139__32FFF970C5653EF40E79FD52038132A8.pdf?sequence=1&isAllowed=y [Con acceso el 10 de mayo de 2020].

Chicote, J. (2018, Septiembre 14). El presidente del Gobierno cometió plagio en su tesis. *ABC*. Disponible en: https://www.abc.es/espana/abci-presidente-gobierno-cometio-plagio-tesis-201809130552_noticia.html [Con acceso el 7 de julio de 2019].

Coulthard, M., y Johnson, A. (2010). *The Routledge handbook of forensic linguistics* [El manual de Routledge sobre lingüística forense]. Milton Park, Abingdon, Oxon: Routledge.

España. Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. Boletín Oficial del Estado, 24 de noviembre de 1995, núm. 281, pp. 33987-34058.

García Rebollar, A. (2010). Notas sobre la diplomacia económica. *Boletín económico del ICE, Información Comercial Española 3.000*, 23-34. Disponible en: https://docplayer.es/86280039-Notas-sobre-diplomacia-economica.html [Con acceso el 10 de mayo de 2020].

Guillén Nieto, V. (21 de noviembre de 2018). "La detección de plagio en textos especializados". Conferencia ofrecida en la Universidad de Murcia.

Iglesias Onofrio, M, y Molina Rabadán, D. (2008). La Estrategia Marca País en la sociedad informacional: los casos de España y Ecuador. *Historia Actual Online*, 16, 109-126. Universidad de Cádiz: Asociación de Historia Actual. Disponible en: https://dialnet.unirioja.es/servlet/articulo?codigo=2719256 [Con acceso el 10 de mayo de 2020].

Johnson, A. y Wright, D. (2014). Identifying idiolect in forensic authorship attribution. *Language and Law / Linguagem e Direito*, 1(1): 37-69. Disponible en: https://www.researchgate.net/publication/299134045_Identifying_idiolect_in_forensic_authorship_attribution_an_n-gram_textbite_approach [Con acceso el 10 de mayo de 2020].

Nini, A. y Grant, T. (2013). Bridging the gap between stylistic and cognitive approaches to authorship analysis using systemic functional linguistics and multidimensional analysis. *The International Journal of Speech, Language and Law*, 20(2): 173-202.

Real Academia Española. (2019). Diccionario de la lengua española, 23ª ed. Disponible en: https://dle.rae.es/ [Con acceso el 10 de mayo de 2020].

Sánchez Castejón, P. (2012). *Innovaciones de la diplomacia económica española: análisis del sector público (2000-2012)*. Disertación doctoral no publicada, Universidad Camino José Cela, Madrid. Disponible en: https://estaticos.elperiodico.com/resources/pdf/8/9/1536923681398.pdf?_ga=2.81391255.1489960738.1564219630-346386610.1557413271 [Con acceso el 10 de mayo de 2020].

Saner, R. y Yiu, L. (2001). International economic diplomacy: mutations in post-modern times. Netherlands Institute of International Relations 'Clingendael', La Haya.

Turnitin. (2019). Historia. Disponible en: https://www.turnitin.com/ [Con acceso el 6 de julio de 2019].

Zechner, M., Muhr, M., Kern, R., y Granitzer, M. (2009). External and intrinsic plagiarism detection using vector space models. Stein, Rosso, Stamatatos, Koppel, Agirre (eds.). *3rd pan workshop: Uncovering plagiarism, authorship and social software misuse*, Universidad Politécnica de Valencia, Editorial UPV, 47-55. Disponible en: https://pdfs.semanticscholar.org/160b/400d726eb042d0867d537c447e858716e7b7.pdf [Con acceso el 10 de mayo de 2020].