El uso de recursos tecnológicos en lingüística forense

SHEILA QUERALT
Directora
Laboratorio SQ-Lingüistas Forenses
C/ Pallars 193, Edificio Spaces
08005 Barcelona (España)
E-mail: sheila.queralt@cllicenciats.cat

EL USO DE RECURSOS TECNOLÓGICOS EN LINGÜÍSTICA FORENSE

RESUMEN: En este capítulo se muestran algunas de las herramientas informáticas disponibles que pueden ayudar a agilizar la labor del lingüista forense en los casos de análisis de muestras escritas, como son la construcción de perfiles sociolingüísticos, la atribución de autoría, la desambiguación de textos o la detección de plagio. También se ofrece una reflexión sobre la figura del perito lingüista y la dimensión ética de su profesión. Finalmente, es imprescindible tener en cuenta que los conocimientos y la experiencia del perito lingüista siguen siendo claves e imprescindibles para que el uso de los avances tecnológicos sea eficiente y que se consigan resultados rigurosos y fundamentados en teorías científicas validadas.

PALABRAS CLAVES: lingüística forense; ambigüedad textual; nivel lingüístico; perfiles sociolingüísticos; atribución de autoría; plagio.

THE USE OF TECHNOLOGICAL RESOURCES IN FORENSIC LINGUISTICS

ABSTRACT: This chapter shows some of the available computer tools that can help quicken the work of forensic linguists in the analysis of written samples for forensic purposes such as sociolinguistic profiling, authorship attribution, text disambiguation or plagiarism detection. It also reflects on the role of the linguistics expert and the ethical dimension of this profession. In conclusion, it is proposed that the knowledge and experience of the linguistics expert are still key and indispensable to use the latest technological tools efficiently and obtain rigorous results grounded on validated scientific theory.

KEY WORDS: forensic linguistics; textual ambiguity; language proficiency; sociolinguistic profiling; authorship attribution; plagiarism.

L'UTILISATION DES RESSOURCES TECHNOLOGIQUES EN LINGUISTIQUE JUDICIAIRE

RÉSUMÉ : Ce chapitre présente quelques-uns des outils informatiques disponibles qui peuvent aider à rationaliser le travail du linguiste légiste dans les cas d'analyse d'échantillons écrits, tels que la construction de profils sociolinguistiques, l'attribution de la paternité, la clarification des textes ou la détection plagiat. Il propose également une réflexion sur la figure de l'expert linguiste et la dimension éthique de son métier. Enfin, il est essentiel de garder à l'esprit que les connaissances et l'expérience de l'expert linguiste restent essentielles et essentielles pour que l'utilisation des avancées technologiques soit efficace et aboutisse à des résultats rigoureux basés sur des théories scientifiques validées.

MOTS CLÉS : linguistique judiciaire ; ambiguïté textuelle ; niveau linguistique ; profilage sociolinguistique ; l'attribution de la paternité ; plagiat.

Fecha de Recepción 10/07/2020
Fecha de Revisión 03/08/2020
Fecha de Aceptación 01/09/2020
Fecha de Publicación 01/12/2020
DOI: http://dx.doi.org/10.25267/Pragmalinguistica.2020.i28.11

El uso de recursos tecnológicos en lingüística forense

SHEILA QUERALT

1. INTRODUCCIÓN

La lingüística forense es la rama de la lingüística aplicada que se encarga de analizar y describir los usos de la lengua que hacen distintos participantes en contextos judiciales e investigativos, por lo que en varias ocasiones también se ha descrito como el punto de encuentro entre la lengua y el derecho. Gibbons y Turell (2008) distinguen tres grandes ámbitos: el lenguaje legal (*the language of the law*), el lenguage judicial (*the language of the court*) y el lenguage evidencial o probatorio (*language as evidence*).

El primero de ellos, el ámbito del lenguaje legal comprende estudios descriptivos de diversos tipos y géneros textuales propios del derecho, cuyas peculiaridades han interesado a juristas y lingüistas durante décadas (*e.g.* Lavery, 1921; Tiersma, 1999; Stygall, 2010), como contratos, sentencias judiciales, leyes y un largo etcétera. Relacionadas con este ámbito, también abundan publicaciones que abogan por el uso de un lenguaje comprensible para el mayor número de usuarios posible en contextos legales y administrativos (Montolío, 2012 o Poblete *et al.*, 2018, por ejemplo).

El segundo ámbito, el del lenguaje judicial, explora los usos lingüísticos que hacen los distintos participantes en los procesos relacionados con la impartición de la justicia (como jueces, magistrados, testigos, abogados, fiscales, agentes de policía, víctimas o sospechosos). Estos procesos incluyen, entre muchos otros, juicios ordinarios, entrevistas policiales o procedimientos abreviados. En este tipo de contextos el lingüista forense analiza el tipo de preguntas, las estrategias discursivas, el vocabulario y la adecuación de todos ellos al contexto y al interlocutor, por nombrar solo algunos de los parámetros.

El tercer ámbito, el del lenguaje evidencial, incluye el gran abanico de situaciones en que se requieren los conocimientos de los lingüistas forenses durante procesos policiales o judiciales, ya sea como asesores o como expertos que aporten informes periciales. Dentro de ese abanico, que sigue creciendo para satisfacer las necesidades de la sociedad en su conjunto, se encuentran los distintos ejemplos que se presentan en el cuerpo de esta comunicación. Las características comunes entre los casos que tratamos son, en primer lugar, el uso de recursos tecnológicos por parte del lingüista forense y, en segundo, la modalidad escrita (y no oral) de las muestras analizadas. Sin embargo, antes de abordar los distintos ejemplos de casos que requieren la intervención de un experto en lingüística forense, el siguiente apartado repasa los métodos utilizados en la disciplina.

2. METODOLOGÍA DE TRABAJO

Independientemente de cuál sea el tipo de encargo que recibe un lingüista forense, la primera pregunta que debe plantearse es si el material que se le ha entregado reúne las condiciones necesarias para ser sometido a los distintos tipos de análisis que se requieran. Para responder a dicha pregunta, el experto debe explorar el material en relación a dos criterios que le servirán para determinar si existe o no caso lingüístico: el criterio de longitud y el de calidad (Queralt, 2014: 37; Garayzábal Heinze, Queralt Estévez y Reigosa Riveiros, 2019: 44).

En cuanto a la longitud de las muestras lingüísticas, el experto debe ser siempre consciente de que, en principio, a mayor cantidad de material analizable (número total de palabras en el caso de la lengua escrita y tiempo de grabación del habla de un individuo en el de la lengua oral), mayor cantidad de datos debería poder extraer de él. Aun así, el experto también debe recordar que el criterio de longitud no es suficiente por sí mismo, sino que está íntimamente relacionado con el criterio de calidad de las muestras. Este segundo criterio establece que, para que un lingüista forense pueda analizar el material que se le ha entregado, debe poder hallar en ese material información útil y relevante.

Para eso, es imprescindible que, en el caso de las muestras orales, la grabación sea de una calidad suficiente como para permitir el análisis lingüístico, tanto por sus características técnicas (derivadas de factores como el equipo que se usó para hacerla, el formato del archivo de audio, la disponibilidad de los programas informáticos necesarios para su procesamiento, etc.) como por características auditivas y perceptivas (como la presencia o ausencia de ruidos, el volumen en que se comunican los hablantes, la distancia que les separa del micrófono, etc.). En el caso de las muestras escritas, los requisitos incluyen la posibilidad de acceder a los documentos (ya sea en formato físico o digital), su legibilidad y su comprensibilidad, entre otros. Además, el criterio de calidad también abarca la calidad lingüística de las muestras orales o escritas. Una muestra tendrá calidad lingüística cuando permita la identificación de características sociolingüísticas que el lingüista forense pueda usar para intentar dar respuesta a la pregunta que se le ha planteado.

Así, por ejemplo, podríamos aplicar los dos criterios a un caso ficticio. El encargo que se le plantease al lingüista forense podría ser intentar concretar el mayor número posible de características sociolingüísticas de un hablante concreto (nivel educativo, edad, origen geográfico, conocimiento de una o múltiples lenguas, profesión, etc.). Imaginemos que la grabación que nos aportan es de una interacción entre un cliente y un sistema automático empleado por una compañía telefónica. En ella, el hablante responde con una única palabra (muchas veces, sí o no) a varias preguntas que se le hacen y que contienen varias alternativas entre las que debe elegir. En un caso así, seguramente podremos llegar a la conclusión de que el material lingüístico no reúne las condiciones necesarias para poder ser analizado, por muy larga

que sea la grabación (criterio de longitud), ya que no se cumpliría el criterio de calidad. Concretamente, porque nuestra grabación ficticia no presentaría características lingüísticas que aportasen información útil ni relevante para determinar características interesantes para quien nos hizo el encargo sobre el hablante en cuestión.

Aun así, se podría llegar a una conclusión opuesta si, en nuestra grabación hipotética, el cliente mostrase características muy peculiares como una pronunciación muy marcada de ciertos fonemas o un trastorno que afectase a su habla de forma singular o si, en vez de contestar siempre de manera escueta, produjese largos turnos de habla (por ejemplo, en nuestra conversación ficticia entre humano y sistema automático, el primero podría disgustarse con su interlocutor y reprenderlo largamente o dirigirse a otra persona para quejarse de su conversación).

Una vez se ha podido determinar que existe caso lingüístico y que el material cumple con los criterios de longitud y de calidad, el siguiente paso que debe tomar el lingüista forense es el diseño del análisis que llevará a cabo. Para diseñar la metodología de análisis más adecuada para responder a su pregunta con el material de que dispone, el experto debe, en primer lugar, poder seleccionar las teorías lingüísticas que le puedan resultar más útiles en su encargo. Así pues, es de vital importancia que el analista tenga unos sólidos conocimientos de lingüística en su sentido más amplio, ya que a mayor manejo de las teorías, conceptos y aportaciones que conforman el conocimiento actual sobre los sistemas lingüísticos y los usos que hacemos de ellos en sociedad, más opciones tendrá el experto a su alcance para afrontar cada trabajo desde la perspectiva que mejor se adapte a unos objetivos específicos.

Además, dicho conocimiento no debe ser exclusivamente teórico, sino que un lingüista forense debe ser capaz de aplicarlo a las muestras con las que trabaja. Asimismo, debe también mantenerse al corriente de los avances, tanto teóricos como metodológicos, que se dan continuamente, en general, en lingüística general y, en particular, en lingüística forense. Una vez ha decidido desde qué perspectiva teórica va a proceder y qué metodología aplicará, el experto debe concretar qué variables lingüísticas le permitirán responder a su incógnita, de qué herramientas deberá servirse para lograrlo y cómo deberá utilizarlas para ser lo más eficiente posible.

El diseño del análisis que se hará puede dar como resultado métodos de tipo cualitativo, cuantitativo o combinado. El primero se caracteriza por el uso de variables no cuantificables, pero que permiten analizar la muestra analizada en profundidad. En cambio, los métodos cuantitativos se basan en la cuantificación de variables y permiten la obtención de valores porcentuales fácilmente comparables entre sí y que pueden ser representados en figuras y gráficos. Finalmente, es posible complementar un tipo de método con el otro, como ocurre en los métodos combinados. Estos conjugan los dos tipos anteriores, de manera que los análisis en que se aplican pueden describir las muestras lingüísticas desde una perspectiva holística y conseguir

una imagen más completa que mediante la aplicación de un solo tipo de método.

Aun así, en lingüística forense, como sucede en otras disciplinas, la aplicación de métodos cualitativos es imprescindible, independientemente de que pueda o no preceder o suceder a métodos cuantitativos. Esto se debe a que los resultados cuantitativos, aunque pueden complementar a los cualitativos, no pueden aportar información suficiente sin que el experto los someta a su interpretación, que necesariamente debe ser cualitativa. De este modo y a pesar de los grandes avances que se ha hecho recientemente en tecnologías automáticas de procesamiento de material lingüístico, el uso de programas automáticos y de medidas cuantitativas no debe ni puede sustituir la valoración cualitativa del lingüista ni servir aisladamente como un análisis suficiente para extraer conclusiones en lingüística forense.

A continuación, se presentan distintos ejemplos de casos en los que se puede observar cómo los lingüistas forenses pueden servirse de ciertos recursos tecnológicos para complementar parte de sus análisis en periciales lingüísticas.

3. ANÁLISIS DE AMBIGÜEDADES TEXTUALES

La ambigüedad se encuentra en muchos niveles lingüísticos, como el semántico, el sintáctico, el de los actos de habla o en el plano de las presuposiciones (Reyes, 2018: 95). Un elemento ambiguo (palabra, oración, enunciado) se puede interpretar de varias formas, lo cual, en ocasiones, puede beneficiar al hablante que lo produce (por ejemplo, en titulares de noticias, cuyo objetivo es resultar llamativos). Así, muchas de las ambigüedades y vaguedades existentes en el lenguaje legal pueden ser intencionales. Esto se debe a los objetivos que deben cumplir ciertos géneros textuales, como las leyes o los contratos, cuya redacción está constreñida por la necesidad de regular situaciones y, a la vez, prever un gran número de circunstancias que pueden ocurrir en la vida real. Además, cabe subrayar que dicha redacción no es baladí, sino que influencia la aplicación de las leyes y cláusulas acordadas por los juristas. No obstante, incluso los textos legales, en que cada oración (o cláusula o artículo o epígrafe…) se somete a muchas revisiones y debe ser consensuada entre varias partes, presentan ambigüedades no intencionales.

Sean o no planeadas, las ambigüedades en textos legales pueden tener mayores consecuencias que las que se dan continuamente en otros discursos debido a la naturaleza de estos géneros textuales. Como avanzábamos, el redactado de las leyes afecta directamente a la interpretación que hacen de ellas los juristas, encargados de hacerlas cumplir, y puede generar debates y discrepancias de gran relevancia social. Muchas de estas disputas se originan por ambigüedades sintácticas (combinaciones de palabras que dan lugar a significados oracionales distintos) o semánticas (palabras con más de un posible significado en su contexto).

En una disputa real, se solicitó la intervención de una lingüista forense para que analizase un fragmento de un convenio colectivo. El elemento que originó la controversia era la conjunción *salvo* en uno de los apartados que describían los colectivos a los que se podía aplicar la cláusula de subrogación: «No obstante lo anterior, quedan excluidos de la aplicación de la presente cláusula de subrogación aquellos empleados/as que sean directivos de su empresa, así como aquellos unidos por vínculos de consanguinidad y afinidad, salvo que acrediten la existencia de relación contractual». Para la empresa, la excepción «salvo que acrediten relación contractual» solo debía aplicarse al elemento inmediatamente anterior («aquellos unidos por vínculos de consanguinidad y afinidad»), es decir, hijos, hermanos o parejas de directivos, pero no a estos últimos. Sin embargo, la defensa alegó que la excepción abarcaba también a los directivos con relación contractual.

En este caso, se diseñó una metodología compuesta por tres análisis: lógico-formal, semántico y sintáctico. El primero tenía como objetivo estudiar la construcción lógica del lenguaje utilizado en el fragmento del convenio en cuestión a partir de la combinación de sus formantes lingüísticos. Al inspeccionar el rol de cada elemento y su contribución a la lógica global del texto, los resultados de este doble análisis permiten acceder a una comprensión holística del fragmento y del texto.

En cuanto al análisis semántico, se observaron los significados de los términos empleados en el fragmento en disputa mediante el uso de dos herramientas. Por un lado, el *Diccionario de la lengua española* (DLE; RAE, 2017) y, por otro, la plataforma *Enclave RAE*[1], que recoge varios servicios lingüísticos. Como se ha explicado, la controversia en este caso se debía a la interpretación de la conjunción *salvo*, que en el DLE se define como «excepto», conjunción de la cual, a su vez, se indica que «Introduce un elemento que supone una excepción dentro de un conjunto o una totalidad que pueden o no estar expresos». (RAE, 2017).

La aplicación de este significado al fragmento analizado muestra que *salvo* introduce una condición (la existencia de relación contractual) que invalida la exclusión de la subrogación en el caso de los directivos y empleados que estén unidos por vínculos de consanguinidad y afinidad. Por tanto, del fragmento se desprende que la exclusión de la subrogación no es aplicable si se demuestra una relación contractual.

El tercer análisis, el sintáctico, tenía como objetivo identificar los distintos componentes de la oración (sujeto, verbo, complementos) y especificar su estructura y las relaciones que vinculan esos componentes. Se llevó a cabo de forma manual y posteriormente se hizo uso de una herramienta informática para su representación gráfica. Los analizadores sintácticos automáticos precisan de un texto de entrada más sencillo y organizado que el original. En este caso, por este motivo, se redujo el texto de entrada manteniendo la estructura principal, pero eliminando las estructuras secundarias y simplificando los nexos de coordinación. En esta ocasión,

[1] Accesible a través de https://enclave.rae.es.

se utilizó el analizador morfosintáctico *Stilus* (*MeaningCloud*, s.f.) para representar una versión simplificada de la cláusula: «Aquellos empleados directivos y aquellos empleados con vínculos de consanguinidad y afinidad quedan excluidos de la cláusula de subrogación, salvo que acrediten la existencia de relación contractual».

El *software* ofrece un árbol sintáctico claro (Figura 1) en el que se puede observar una única oración compleja, con un sujeto compuesto por dos sintagmas coordinados por una conjunción copulativa, un predicado principal unido a una oración subordinada mediante la conjunción *salvo que*.

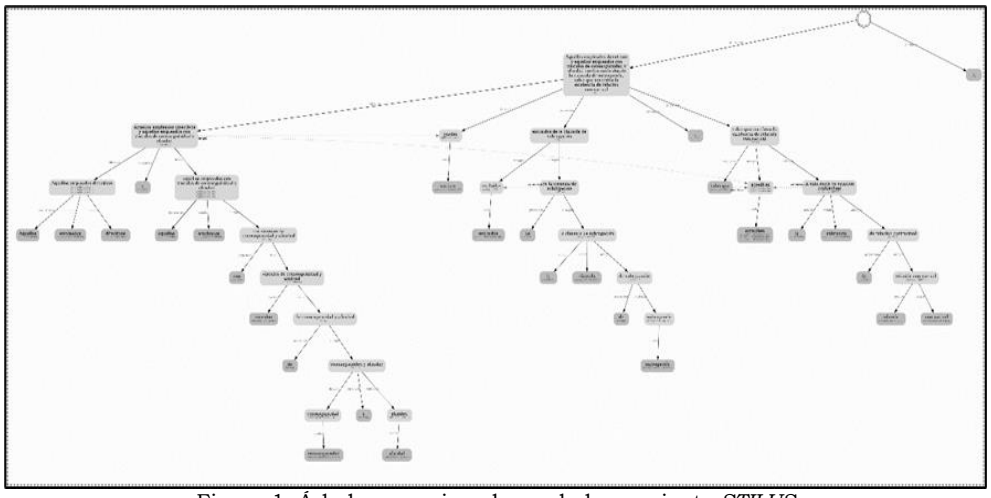


Figura 1: Árbol proporcionado por la herramienta *STILUS*

En este fragmento son de especial interés los constituyentes coordinados y subordinados. La coordinación de componentes indica que las unidades lingüísticas contiguas tienen relación, comparten la misma función y jerarquía sintácticas y están relacionadas también semánticamente. La subordinación de elementos implica una dependencia a otros componentes oracionales a los que complementan o modifican.

En el fragmento analizado, la relación de coordinación mediante la locución conjuntiva, *así como* indica que «aquellos empleados/as que sean directivos de su empresa» y «aquellos unidos por vínculos de consanguinidad y afinidad» comparten la función sintáctica de sujeto y poseen la misma jerarquía, por lo que ambos grupos de empleados reciben la acción del verbo *quedan excluidos*. Además, la conjunción *salvo* introduce una oración subordinada condicional. Por tanto, la relación sintáctica de subordinación permite concluir que la oración condicional exceptiva se aplica a todo el sujeto. Por otro lado, el orden en el que se presenta la condición es el canónico, como se indica en la *Nueva gramática de la lengua española* (2009: 3583): «tienden a ir pospuestas ("apódosis-prótasis") las prótasis condicionales focalizadas, como las encabezadas por, *sobre todo, al menos, salvo o excepto*».

Teniendo en cuenta los resultados de los tres análisis realizados, la conclusión que defendió el informe pericial aportado fue que el fragmento transmitía la siguiente proposición condicional: «si los empleados directivos y los empleados con vínculos de consanguinidad y afinidad acreditan una relación contractual, deben ser subrogados». Por tanto, según la lingüista forense, la subrogación debía practicarse en ambos casos siempre que se acreditase la existencia de contratos. Varias sentencias judiciales (véase por ejemplo la sentencia 00207/2018 del Juzgado de lo Social Nº2 de Badajoz) expedidas por esta disputa han desestimado la interpretación contraria defendida por la empresa (Soriano, 2018).

4. ANÁLISIS DEL NIVEL LINGÜÍSTICO EN PRUEBAS SELECTIVAS

Durante procesos de selección o renovación de su personal, las empresas pueden evaluar las distintas competencias de sus trabajadores o candidatos. En este tipo de situaciones, los servicios en lingüística forense pueden ser útiles cuando las competencias que se evalúan son lingüísticas, es decir, cuando interesa medir el nivel de conocimientos de lengua de los examinados.

Dichos procesos de evaluación pueden ser el objeto de disputas cuando hay sospechas de que el nivel de dificultad de las pruebas no se corresponde con el especificado en las bases de la convocatoria. Un lingüista forense experto en la lengua en cuestión y con experiencia docente como profesor de esa lengua puede analizar las preguntas o instrucciones del examen para determinar si se ajusta a los estándares de distintos niveles de competencia lingüística existentes.

En este caso, se examinaba por escrito el nivel de inglés «general y técnico» de los trabajadores de una gran empresa. El test se componía de unas cincuenta preguntas con cuatro posibles respuestas, de las que solo una era correcta.

La pregunta que debía responder el informe solicitado a las lingüistas forenses que participaron como expertas en el caso era si el nivel evaluado por ese test era igual, inferior o superior al nivel B1 del Marco Común Europeo de Referencia para las Lenguas (MCER, MECD, 2002). Para responderla, se aplicó una combinación de métodos cualitativos y cuantitativos.

El análisis cualitativo exploró la complejidad de las estructuras gramaticales y del vocabulario que debía conocerse para superar el test. El análisis cuantitativo se centró en la identificación y la cuantificación de las preguntas que requerían un nivel superior al especificado en la convocatoria. Para respaldar la identificación de dichas preguntas se utilizaron varias herramientas informáticas.

En primer lugar, se recurrió al servicio *English Vocabulary Profile*[2], desa-

[2] Accesible a través de http://www.englishprofile.org

rrollado por Cambridge University Press en colaboración con otros organismos y universidades[3]. *English Profile* se basa en un corpus de referencia con miles de muestras de alumnos reales provenientes de multitud de países y, por lo tanto, la metodología que utiliza es empírica, ya que da evidencia concreta sobre qué estructuras lingüísticas conocen los aprendices de inglés de todo el mundo en cada nivel del MCER. De este modo, los resultados de *English Profile* se pueden referenciar con mucha más certeza que otras herramientas similares desarrolladas anteriormente (Harrison y Barker, 2015: 4).

Además, es un recurso diseñado para facilitar la comprensión del MCER por parte de profesores y educadores, ya que describe qué aspectos del inglés se suelen aprender en cada nivel. Así pues, puede usarse como una base de datos donde encontrar qué elementos de vocabulario es razonable y adecuado evaluar en cada nivel de aprendizaje. Knight (2015: 98) especifica que permite verificar los puntos de vocabulario y de gramática de ejercicios o exámenes para asegurarse de que sean apropiados para el nivel que se está evaluando. Para cada expresión, se indica el nivel del MCER en que se ha clasificado, su significado en inglés, un ejemplo extraído de un diccionario y otro del corpus de producciones de aprendices de inglés como lengua extranjera. Como se puede observar en la figura siguiente, aparece la expresión buscada, el nivel al cual se clasifica (en este caso, C2), el significado, el ejemplo que aporta el diccionario, y el ejemplo que aporta el estudiante.

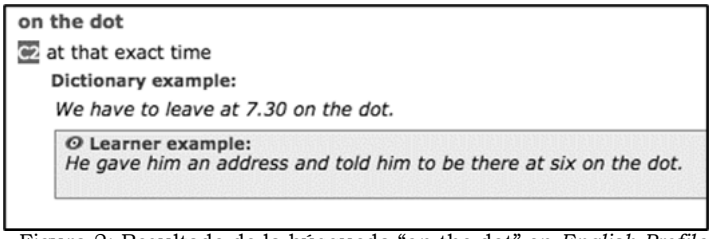


Figura 2: Resultado de la búsqueda "on the dot" en *English Profile*

Después del análisis cualitativo de las preguntas del test, que permitió localizar los elementos léxicos clave que el examinado debía conocer para dar con la respuesta correcta, se aplicó esta herramienta para cuantificar el vocabulario clave del examen por niveles. Así, por ejemplo, en una pregunta en que se debiera completar la oración «We are carrying out a (*respuesta*) on customers' satisfaction with our products» y cuyas posibles respuestas fueran *question*, *profile*, *survey* y *questionnaire*, se clasificaría los ítems *survey* y *profile* como esperables para un nivel B2.

Los resultados de este análisis se completaron mediante el uso de otra de las herramientas disponibles en *English Profile* llamada *Text Inspector*

[3]Cambridge University Press, Cambridge English Language Assessment, University of Cambridge, University of Bedforshire, British Council, English UK, además de socios contribuyentes de datos y otras organizaciones que participan en el proyecto *English Profile Network*, financiado por la Unión Europea (English Profile, 2015).

(*WebLingua*, s.f.). Este recurso permite calcular la dificultad del texto introducido. El texto es dividido en unidades léxicas que se clasifican por nivel según sus distintos significados. De este modo, se recupera una relación de los ítems léxicos del texto cuantificados y clasificados por niveles, que puede visualizarse en forma de gráfico. Las distintas preguntas del test se introdujeron como texto en esta herramienta para contrastar los resultados del análisis del vocabulario.

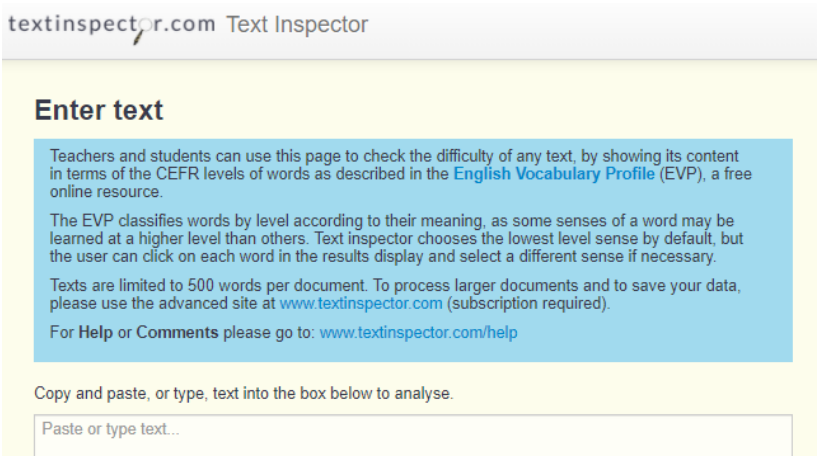


Figura 3: Texto de entrada en *Text Inspector*

Se introdujeron los enunciados de las distintas preguntas del test con sus correspondientes posibles respuestas en la interfaz que muestra la Figura 3, que permite escribir o copiar textos de hasta 500 palabras.

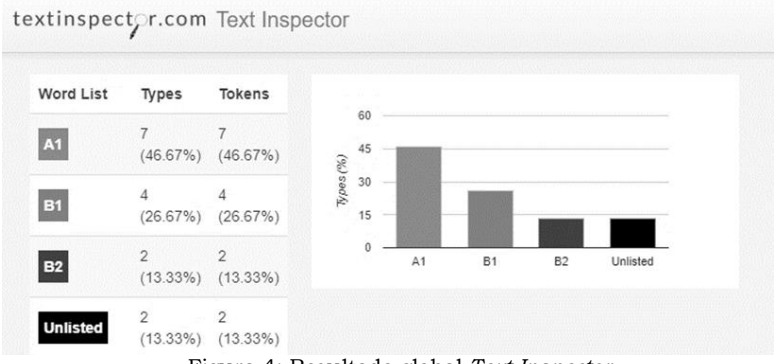


Figura 4: Resultado global *Text Inspector*

Los resultados globales de la clasificación automática que produce esta herramienta tienen el aspecto que muestra la Figura 4. En ella, se observa un gráfico de barras que refleja las cantidades de ítems léxicos pertenecientes a los distintos niveles identificados en el texto introducido. A la izquierda, la leyenda muestra los valores brutos y porcentuales de cada

grupo de elementos, incluyendo aquellos que la herramienta no ha podido clasificar. Además, como muestra la Figura 5, se proporciona la codificación que se ha hecho de cada palabra en el texto introducido. En esta visualización, el usuario puede presionar sobre cada palabra para ver las distintas acepciones vinculadas a cada ítem con su correspondiente nivel del MCER y, si fuera necesario, seleccionar la que se ajusta al significado de esa palabra en el contexto.

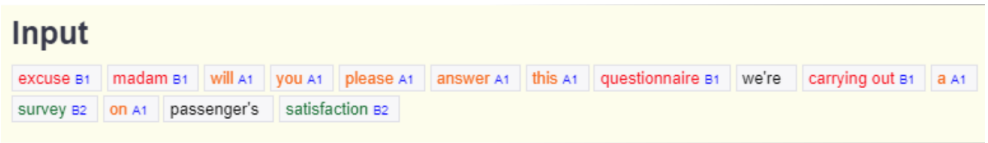


Figura 5: Resultado *Text Inspector* por palabra

Se debe destacar una de las limitaciones de *Text Inspector*. En el caso de que la palabra posea distintos significados y que los significados correspondan a niveles distintos, *Text Inspector* únicamente refleja el nivel de uno de los significados, el más bajo. De este modo, es altamente recomendable revisar individualmente cada una de las palabras para confirmar que efectivamente el programa arroja el nivel del significado correcto de la palabra introducida y, además, valerse de la herramienta *English Vocabulary Profile* que sí discierne entre el nivel de los distintos significados, como se observa en la Figura 6.

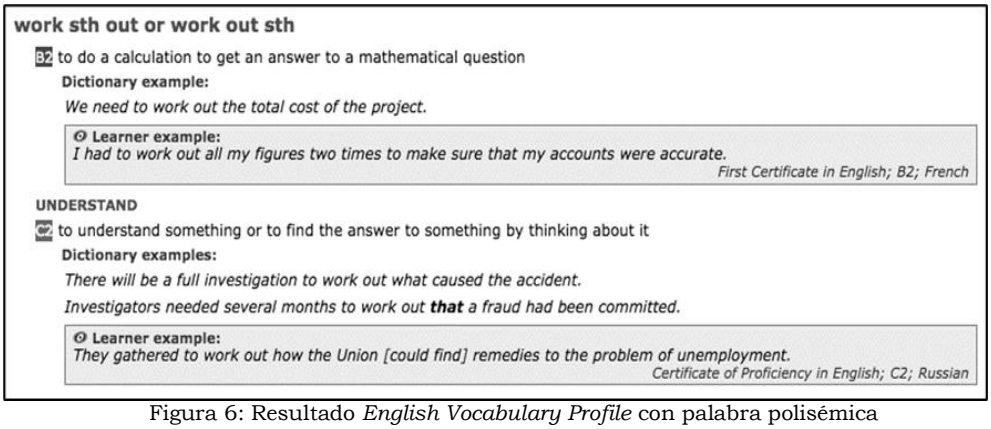


Figura 6: Resultado *English Vocabulary Profile* con palabra polisémica

En este caso, la expresión buscada (*work something out*) tiene dos significados y cada uno de ellos es esperable en un nivel de conocimiento lingüístico distinto: el primero, equivalente a "calcular" es esperable en un nivel B2, mientras que el segundo significado, equivalente a "comprender" no es esperable hasta el nivel C2.

Los análisis llevados a cabo en este caso permitieron calcular qué porcentajes del total de preguntas del test correspondían a cada nivel de com-

petencia. Concretamente, se pudo concluir que para responder correctamente a la mayoría de las preguntas del test los examinados debían poseer un nivel de competencia de inglés como lengua extranjera superior al B1 del MCER.

5. Elaboración de perfiles sociolingüísticos

Uno de los axiomas más ampliamente aceptados en lingüística forense es que las producciones lingüísticas de los hablantes ya sean orales o escritas, reflejan características vinculadas a su identidad como miembros de distintos colectivos sociales. Esas características lingüísticas difieren de las producciones de un individuo a las de otro (variación entre hablantes), de modo que es posible postular la existencia de «estilos idiolectales» únicos (Turell, 2010). Este tipo de variación lingüística es mayor que la variación entre las producciones de un mismo hablante (intra-hablante), por lo que es posible identificar marcadores lingüísticos en producciones de autor desconocido para obtener información sociolingüística. Esta tarea se conoce en lingüística forense como elaboración de perfiles sociolingüísticos y puede encargarse como parte de investigaciones policiales o privadas con el fin de acotar el número de posibles autores de la muestra o para reconducir la búsqueda de sospechosos. Suele preceder a tareas de comparación de muestras lingüísticas en fases investigativas en las que solo se dispone de un conjunto de producciones.

La elaboración de perfiles sociolingüísticos se sustenta en los principios teóricos de que cada muestra lingüística es única e irrepetible, y contiene marcadores lingüísticos propios de su autor. Además, comúnmente (y especialmente en contextos informales), los autores y los receptores de las muestras lingüísticas no fijan su atención en sus características ni en la forma en cómo se producen, y no son conscientes de los marcadores sociolingüísticos en ellas, de manera que no son detectados por hablantes que podrían tener la intención de falsearlos o imitarlos. Aun así, en casos en que el autor intentase manipular sus usos lingüísticos, el experto observaría inconsistencias en el uso de los marcadores identificados (Queralt y Giménez, 2018).

Los marcadores sociolingüísticos identificados en las muestras pueden aportar información de diversas características sociales e individuales del autor como su profesión, rango de edad, sexo, nivel educativo, creencias religiosas, sesgos ideológicos, etc. La investigación sociolingüística ha demostrado a lo largo de las décadas que hombres y mujeres tienden a mostrar comportamientos lingüísticos que siguen patrones distintos (ver Coates, 2004, especialmente el segundo capítulo). Asimismo, un gran número de estudios se han centrado en las características lingüísticas de varias generaciones de hablantes y existe consenso entre lingüistas sobre la tendencia de hablantes pertenecientes a distintos grupos de edad a desarrollar y mantener usos gramaticales y léxicos que les unen a otros

miembros de su propio grupo etario y les separan de otros (*e.g.* Silva-Corvalán 2001: 101-103).

En cuanto a la profesión, el nivel educativo o las creencias religiosas e ideológicas pueden observarse, por ejemplo, en el uso o la ausencia de términos especializados, diferentes estilos expresivos, gramática y puntuación estándares, figuras retóricas (sobre todo metáforas) asociadas a ciertos tipos textuales especializados o elementos propios de ciertos registros o géneros textuales. De forma parecida, otros rasgos socio-colectivos del autor que pueden detectarse en sus usos lingüísticos incluyen su origen geográfico (debido a la existencia de geolectos), su origen étnico o cultural, si se trata o no de un hablante nativo de la lengua utilizada o si puede estar expuesto o tener conocimientos de otras lenguas.

Como parte de una investigación por delitos contra el honor, se requirió la intervención de un lingüista forense para analizar dos mensajes publicados por un usuario anónimo en un foro de opinión por internet. Los mensajes se reproducen en la Tabla 1, aunque se advierte que por razones de confidencialidad se han modificado. El objetivo del encargo era intentar extraer información sociolingüística del posible autor de las publicaciones.

Mensaje 1	Mensaje 2
UFF, qué alivio!!!! Por fin [cadena de televisión] ha suprimido uno de los programas que más manchaba su reputación, ahora que quiten [programa de televisión] y casi no habrá telebasura. Qué dice el otro de echarles de menos?? En fin, solo gentuza sin mente echaría de menos esta mierda.	Pero quien se ha creído que es este tío???? Este tipejo es escoria y la gentuza que le aplaude, el [presentador de televisión] o demás morralla son lo peor de lo peor!!!! [otro usuario] lo puede haber dicho de manera políticamente incorrecta pero describe a la perfección lo que pasa en esa cadena de mierda llena de drogatas

Tabla 1: Corpus de análisis

En esta ocasión, la metodología empleada consistió en un análisis cualitativo preliminar con el fin de identificar marcadores sociolingüísticos potencialmente útiles. Algunos de los elementos identificados en este primer análisis fueron léxicos, como *telebasura*, *gentuza* o *echar de menos* en el primer mensaje y *tipejo*, *ser escoria*, *gentuza* o *morralla* en el segundo. Los marcadores identificados se sometieron a un segundo análisis que pretendía contrastarlos con corpus de referencia mediante el uso de herramientas informáticas para obtener información acerca de las características geográficas más probables del autor anónimo.

Uno de los corpus utilizados fue la sección Web/Dialects de *El corpus del español*[4], que recoge muestras de 21 países de habla hispana. Como refleja

[4] Creado por Mark Davies de Brigham Young University en 2016, contiene 2.000 millones de palabras procedentes de unos 2 millones de páginas web de más de 20 países de habla hispana (https://www.corpusdelespanol.org). Accesible a través de https://www.corpusdelespanol.org/web-dial/.

la Figura 7, para *telebasura*, por ejemplo, el corpus mostró un total de 933 resultados distribuidos de la siguiente manera: 609 resultados para España, 83 para Estados Unidos, 74 para México, 31 para Colombia, 28 para Ecuador, 23 para Perú, y menos de 15 resultados para el resto de los países.

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
GENERAL	306	935.4	0.33	
BLOG	627	1,049.8	0.60	
México	74	246.0	0.30	
Guatemala	1	54.3	0.02	
El Salvador	11	36.5	0.30	
Honduras	2	35.1	0.06	
Nicaragua	1	32.4	0.03	
Costa Rica	8	29.6	0.27	
Pánama	6	22.3	0.27	
Puerto Rico	3	32.2	0.09	
Rep Dom	3	33.7	0.09	
Cuba	10	63.2	0.16	
Venezuela	4	98.2	0.04	
Colombia	31	166.5	0.19	
Ecuador	28	52.4	0.53	
Bolivia	3	39.4	0.08	
Perú	23	107.3	0.21	
Chile	10	66.2	0.15	
Paraguay	4	29.8	0.13	
Uruguay	10	38.7	0.26	
Argentina	9	169.4	0.05	
España	609	426.6	1.43	
EEUU	83	166.0	0.50	
TOTAL	933			SEE ALL TOKENS

Figura 7: Resultados de 'telebasura' en Web/Dialects de *El corpus del español*

Como se puede observar, la diferencia entre España y el resto de los países es muy notable, lo cual debe interpretarse teniendo en cuenta que el corpus recoge un número diferente de palabras para cada país (que oscila entre 24.698.769 para Panamá y 459.312.821 para España), pero aun así señala un uso mucho más frecuente en España que en los demás. Esto puede comprobarse fijándonos en los resultados en cuanto a número de ocurrencias por millón de palabras en vez de en las frecuencias brutas (los 933 resultados para España se traducen a 1,43 y los 83 para Estados Unidos a 0,50 ocurrencias por millón de palabras). Se obtuvieron resultados que igualmente señalaban a variedades utilizadas en España para vocablos como *gentuza* o la combinación *es* + *escoria*: 1.768 resultados para España (4,14 por millón de palabras), seguidos de 357 para Estados Unidos (2,15), en el primer caso y 16 resultados en España (0,04), seguidos de 9 en México (0,04) y 2 en Guatemala (0,04) y en Colombia (0,01) en el segundo caso.

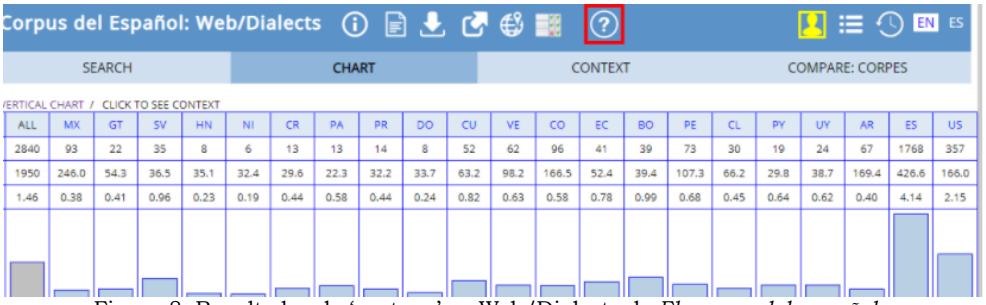
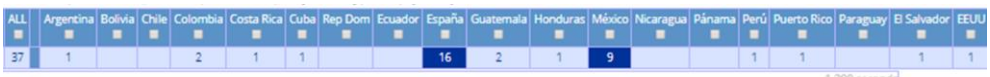


Figura 8: Resultados de 'gentuza' en Web/Dialects de *El corpus del español*



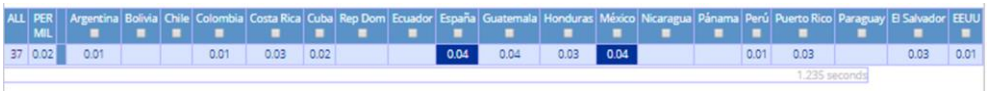


Figura 9: Resultados de 'es + escoria' en Web/Dialects de *El corpus del español*

Si nos fijamos en el número de ocurrencias de *es + escoria* por millón de palabras, observamos que hay una coincidencia en los tres primeros países que parece indicar una probabilidad parecida de que hablantes de cada una de esas variedades utilice la construcción *es escoria*. Si ese fuera el caso, los resultados obtenidos para esta expresión contradicen los resultados mostrados por los demás ítems analizados. En este sentido, cabe recordar que detectar posibles inconsistencias es esperable durante el análisis de muestras lingüísticas, ya que responde a la variación intra- e inter-autor anteriormente mencionadas, es decir, diferentes producciones del mismo hablante presentarán conjuntos distintos de características y, a la vez, no todos los hablantes de una variedad lingüística concreta producirán muestras con exactamente las mismas características. Es por eso por lo que los lingüistas forenses deben trabajar con el mayor número posible de variables teniendo en cuenta las limitaciones prácticas a que son sometidos (tiempo por encargo, carga de trabajo, viabilidad del análisis, etc.). El número de variables analizadas afecta directamente a la solidez de las conclusiones a las que se pueda llegar.

Además de Web/Dialects, otra sección de *El corpus del español* que puede utilizarse en la elaboración de perfiles sociolingüísticos y, concretamente, en casos en los que no se dispone de información sobre la fecha de producción de las muestras en cuestión, es *News on the Web* (NOW). Esta herramienta recoge revistas y diarios publicados a partir de 2012 y contiene más de 5.500 millones de palabras. Además, incorpora unos 170 millones de palabras nuevas cada mes, de forma que puede usarse como un corpus de referencia para medir la variación lingüística más reciente y casi a tiempo real. Así pues, por sus características, esta herramienta puede resultarle muy útil a

un lingüista forense que deba extraer información temporal de un mensaje escrito a mano (y que por tanto no presenta metadatos informáticos), por ejemplo.

6. Comparación de muestras escritas

En varios tipos de casos en que participa, el lingüista forense se enfrenta a dos conjuntos de producciones lingüísticas que debe analizar en busca de similitudes y diferencias. El primero de los conjuntos se compone por muestras de autor desconocido (es decir, dubitadas), mientras que las del segundo son obra de un autor conocido (y, por tanto, se denominan muestras indubitadas). En estos casos, el lingüista empieza por identificar características lingüísticas distintivas en ambos conjuntos. A continuación, compara las características que ha podido identificar y evalúa el grado de similitud o distancia entre las muestras dubitadas e indubitadas. Dicho procedimiento se aplica, entre otros, a muestras de las que se sospecha la originalidad (casos agrupados bajo la etiqueta *detección de plagio*) o la identidad del autor (casos de atribución de autoría).

6.1. Detección de plagio

Solo en la última década, varios casos de posible plagio han tenido una gran repercusión mediática a distintos niveles. Por ejemplo, los supuestos plagios de un discurso de la primera dama de Estados Unidos en 2016 (*Turnitin*, 2016), de la tesis doctoral de un ministro alemán en 2011 (*BBC News*, 2011), o incluso el del himno del Mundial de Fútbol de 2010 (Garrido, 2018). Sin embargo, es razonable pensar que la proporción de plagios que se dan a conocer respecto a los que ocurren diariamente es muy reducida. Un estudio en Portugal, por ejemplo, mostró que un 68,8% de los profesores de distintos niveles educativos encuestados creían que se detectaba menos del 50% de los casos reales de plagio (Dias y Bastos, 2014). Además, las consecuencias para plagiarios probados también varían según distintos factores como la jurisprudencia encargada de juzgarlos o el ámbito social en que se encuentren (entre otras, se pueden dar penas de prisión, dimisiones forzosas, sanciones económicas o consecuencias casi anecdóticas). Todo ello puede interpretarse como un reflejo del poco consenso existente en cuanto a la definición y a la evaluación que se hace del fenómeno del plagio.

Sin embargo, no hay duda de que este es un fenómeno presente en un gran número de ámbitos sociales, ya que se ha denunciado en distintas esferas, como el mundo académico, musical, literario, político, etc. Recientemente, han surgido iniciativas dentro de algunos organismos, como la Universidad Pública de Navarra (Aunión, 2013), que han estimulado el debate sobre las medidas que se pueden implantar para hacer frente a la aparente omnipresencia de prácticas plagiarias. Se han propuesto distintos tipos de medidas, que incluyen, entre otras, las preventivas, las sancionadoras y las

de detección (Montero, 2016). En cuanto a las últimas, es cierto que numerosas instituciones educativas, por ejemplo, ya cuentan con programas informáticos diseñados para detectar coincidencias entre documentos. No obstante, es necesario subrayar las limitaciones de este tipo de aplicaciones, como su capacidad reducida de identificar fragmentos en que el plagiario recurre a estrategias distintas al simple uso de material lingüístico ajeno, como la ampliación o reducción, el uso de sinónimos o el parafraseo, entre otras.

Algunos de los programas más utilizados son, por un lado, Turnitin, Urkund o Viper (este último es gratuito y de prestaciones más reducidas que los anteriores), de uso común en entornos académicos para detectar coincidencias con bases de datos propias o con internet; y, por el otro, CopyCatch Gold, diseñado específicamente para casos de lingüística forense en que se dispone de los conjuntos de muestras dubitadas e indubitadas. En la Figura 10 se observa el resultado global de una comparación mediante CopyCatch Gold. Si hay coincidencia, las frases y palabras aparecen en rojo, mientras que el texto no coincidente aparece en negro. En la Figura 10 se observa en la columna de la izquierda el texto dubitado, como texto base, y el texto indubitado (u original) en la columna de la derecha, como texto comparado. **[P2 S1]** significa que la primera frase del párrafo 2 del texto **TD** coincide con la primera frase del párrafo 2 de TI {**P2 S1**}.

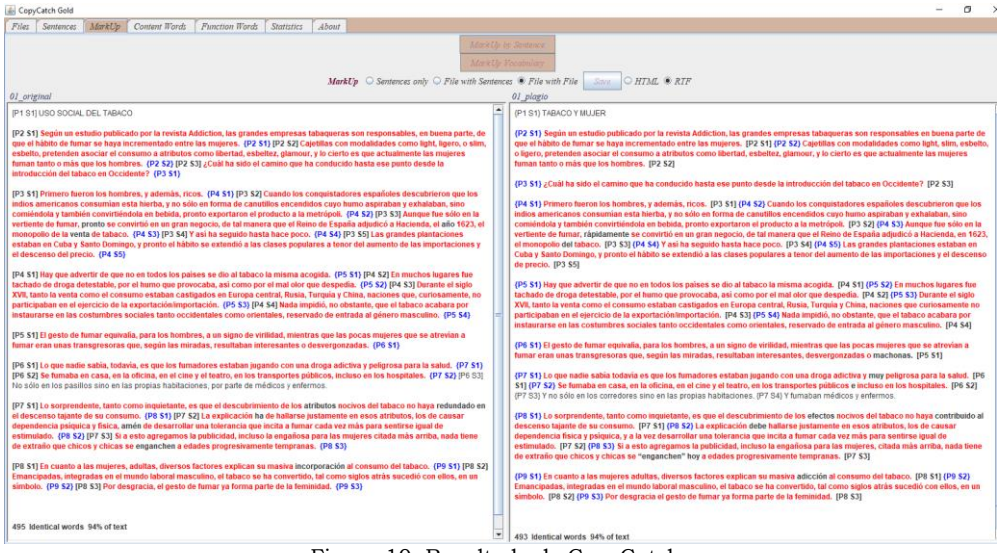


Figura 10: Resultado de CopyCatch

En la Figura 11 se refleja el resultado de Turnitin, el que se indica el porcentaje global de coincidencia (en este caso 44 % con otras fuentes) y se marcan los fragmentos coincidentes indicando con un índice la posible fuente y el porcentaje que supone.

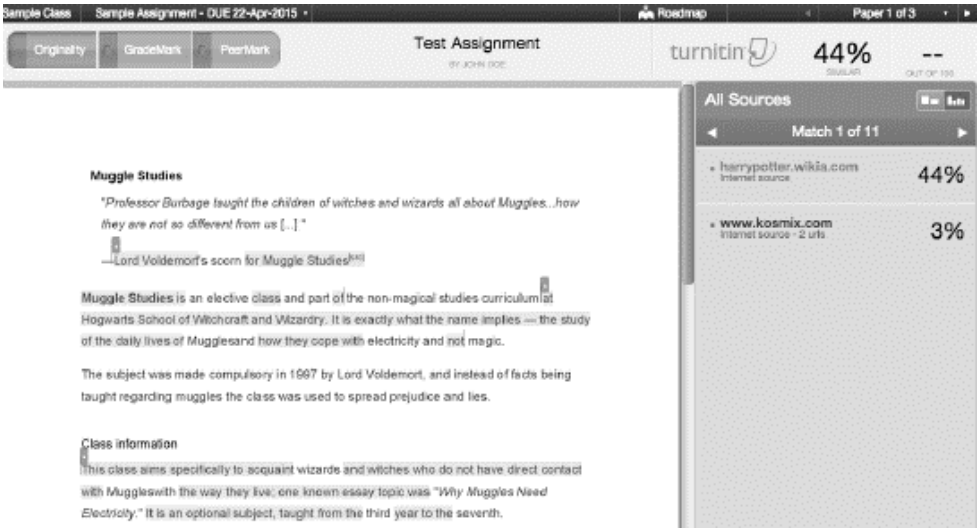


Figura 11: Resultado de Turnitin
Fuente: https://guides.turnitin.com/01_Manuals_and_Guides/Instructor_Guides/Turnitin_Classic_(Deprecated)/21_The_Similarity_Report/Viewing_the_Similarity_Report

El uso de internet como una gran base de datos presenta limitaciones que deben añadirse a las ya mencionadas. La principal es que, a pesar del gran volumen de documentos que sí se encuentran en la red y, por tanto, se escanearán en busca de fragmentos coincidentes, gran parte del material académico, como manuales o libros de texto, no se encuentran digitalizados o no pueden accederse libremente a través de internet. Esto conforma un obstáculo que puede reducir significativamente la efectividad de los programas de detección automática de coincidencias, ya que son precisamente de este tipo de textos de los que los académicos suelen extraer el material que puede constituir un plagio.

Desde una perspectiva lingüística, se pueden distinguir varias estrategias potencialmente constitutivas de plagio. Según Queralt, Marquina y Giménez (2018: 1561-2) y Turell (2011: 72), el plagio lingüístico (no confundir con la copia de ideas) se produce cuando se utilizan las mismas palabras y estructuras gramaticales que la obra fuente para describir ideas (sean o no ajenas), cuando se recurre a la paráfrasis, cuando se usan varias palabras y oraciones sin citar la fuente (incluso cuando haya modificaciones), cuando se mantiene la sintaxis original y se emplean sinónimos o cuando, aun reconociendo el documento original, los cambios introducidos son mínimos (como una o dos palabras, el orden oracional, la voz activa o pasiva, el tiempo o el aspecto verbal).

Un caso que puede ejemplificar la gran variedad de situaciones en que se puede dar plagio y en que, por tanto, pueden intervenir lingüistas forenses es el del ciberataque a escala mundial que se perpetró mediante el programa malicioso WannaCry (Queralt, 2020). El ataque tuvo lugar el 12 de mayo de 2017, entre las 8:00 y las 17:08 horas (UTC) e impactó sobre todo a grandes

empresas de unos 150 países (BBC News 2017), como las españolas Telefónica, Iberdrola o Gas Natural, pero también a otros organismos como el servicio de salud de Reino Unido o el Ministerio del Interior de Rusia (Morente, 2017).

Como se demostró en este caso, el lenguaje informático también puede someterse a análisis lingüísticos. La investigación sobre el ataque con WannaCry reveló que el código del programa se había traducido a 28 lenguas y un estudio de dichas traducciones pudo comprobar que el idioma de origen había sido el inglés (Flashpoint, 2017). Además, el estudio mostró que las traducciones se habían obtenido mediante el servicio gratuito de traducción automática Google Translate.

6.2. ATRIBUCIÓN DE AUTORÍA

El último tipo de encargo que tratamos en esta comunicación es aquel en que se debe resolver una tarea de comparación forense de textos escritos para evaluar la posibilidad de que los distintos conjuntos de textos analizados tuvieran el mismo autor. Siguiendo a Turell (2011: 77), distinguimos entre casos en que se compara el conjunto indubitado con muestras de un único posible autor (atribución de autoría) y otros en se compara con muestras de varios posibles autores (determinación de autoría).

La tipología de casos forenses en que puede ser necesario atribuir la autoría de textos de autor desconocido es muy amplia y comprende, por ejemplo, suicidios sospechosos en que se ha encontrado una nota de despedida que podría haber sido redactada por alguien diferente a la víctima, desapariciones en que se reciben mensajes de la supuesta víctima o de los supuestos secuestradores o casos de extorsión, acoso o usurpación de identidad en que se dispone de un sospechoso.

Además de las consideraciones previas que deben llevarse a cabo antes de aceptar cualquier tipo de encargo en lingüística forense (presentadas en el apartado 2), en atribución de autoría es de especial relevancia evaluar la cualidad y la cantidad de las muestras indubitadas disponibles. El experto debe cerciorarse de que ambos conjuntos puedan compararse en busca de características comunes o dispares. Así, si en un caso de posible suicidio, por ejemplo, el texto dubitado es un mensaje de texto, pero se presenta como conjunto de textos indubitados un total de tres instancias genéricas escritas por el principal sospechoso de haber falsificado el mensaje de despedida, el experto deberá ser muy cauteloso en su análisis preliminar de los conjuntos para poder determinar si es posible llevar a cabo una comparación entre ambos, ya que se trata de géneros textuales muy diferentes.

Según las características del encargo y de la naturaleza de las muestras, un análisis cualitativo de atribución de autoría podría observar distintas variables léxicas (como presencia o ausencia de léxico especializado del campo de la medicina, por ejemplo), morfosintácticas (como presencia o ausencia de concordancia gramatical entre artículos y sustantivos), de puntuación (como uso de comillas latinas, inglesas o simples) o de estructura

textual (como presencia o ausencia de salutación o despedida). Por otro lado, análisis cuantitativos podrían tener en cuenta variables de complejidad (por ejemplo, longitud de párrafo o longitud de frase) o la frecuencia de distintos n-gramas de etiquetas morfosintácticas (secuencias de dos o más elementos lingüísticos etiquetados según su categoría morfológica).

Entre las variables que han mostrado ser más eficaces en tareas de atribución de autoría se encuentran las secuencias de categorías sintácticas o, en inglés, *Part-Of-Speech n-grams* (Turell, 2010; Queralt y Turell, 2013). A pesar de la desventaja de la complejidad del análisis, que requiere un etiquetaje automático y una revisión manual muy exhaustiva, también se trata de una variable muy compleja de modificar voluntariamente por los autores, de manera que representa mejor su estilo idiolectal que otras (Cicres y Queralt, 2019).

Actualmente existen varios programas informáticos que pueden ayudarnos en el etiquetaje de las variables por su categoría sintáctica para su posterior procesamiento estadístico. Una de ellas es *HectorWWW*[5], un analizador morfosintáctico y desambiguador desarrollado por el Institut Universitari de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra. En la Figura 12 se muestra un ejemplo de un texto etiquetado morfosintácticamente con esta herramienta.

```
    |---+----1----+----2----+----3----+----4----+----5----+----6----+----7-
  1 ##      TAG     <div1>
  2 ##      TAG     <p>
  3 ##      TAG     <s>
  4 1       TOK     Buenos  BOS     bueno\JQ--MP
  5 2       TOK     d\xEDas         d\xEDas\N5-MP
  6 ---     DLD     :       EOS     =\DELIM
  7 ##      TAG     </s>
  8 ##      TAG     </p>
  9 ##      TAG     <p>
 10 ##      TAG     <s>
 11 3       TOK     Tengo   BOS     tener\VDR1S-
 12 4       TOK     dudas           duda\N5-FP
 13 5       TOK     de              de\P
 14 6       TOK     que             que\RR---66
 15 7       TOK     proceda         proceder\VJR6S-
 16 8       TOK     presentar               presentar\VI----
 17 9       TOK     ya              ya\D
 18 10      TOK     dicho           decir\VC--SM
 19 11      TOK     escrito         escrito\N5-MS
 20 12      TOK     por             por\P
 21 13      TOK     lo              lo\ANS
 22 14      TOK     siguiente               siguiente\JQ--6S
 23 ---     DLD     :       EOS     =\DELIM
 24 ##      TAG     </s>
 25 ##      TAG     </p>
```

Figura 12: Texto etiquetado morfosintácticamente

En la figura se observa, por ejemplo, el etiquetado del enunciado «Tengo dudas de». Los n-gramas que se extraerían de este enunciado, como se ob-

[5] Accesible a través de http://eines.iula.upf.edu/cgi-bin/hectorwww/hectormain.pl.

serva en la Figura 13, son: un bigrama (secuencia de dos categorías gramaticales) correspondiente a Verbo Indicativo Presente Primera persona del Singular y Nombre Común Femenino Plural; y un trigrama (secuencia de tres categorías gramaticales) compuesto por Verbo Indicativo Presente Primera persona del Singular, Nombre Común Femenino Plural y Preposición.

Tengo	**dudas**	**de**
VDR1S-	N5-FP	P
BIGRAMA		
TRIGRAMA		

Figura 13: Secuencia etiquetada morfosintácticamente

Una vez etiquetado todo el texto se procede al análisis estadístico de los distintos n-gramas. Por lo general, se tiende a utilizar el análisis discri-minante, un análisis estadístico multivariante que permite determinar si hay diferencias entre los escritos de los distintos sospechosos, qué categorías gramaticales los diferencian y atribuir textos anónimos. En la Figura 13 se muestra el resultado gráfico de un análisis discriminante en con 5 textos anónimos (cada uno de ellos representado por una cruz) y cuatro sospechosos con cinco textos cada uno. El candidato más probable de haber producido los textos, según el análisis discriminante, es el sospechoso 4, puesto los textos anónimos se sitúan en la misma zona del gráfico que los de este autor, lo que indica que tiene unas preferencias de uso muy similares de las mismas categorías gramaticales.

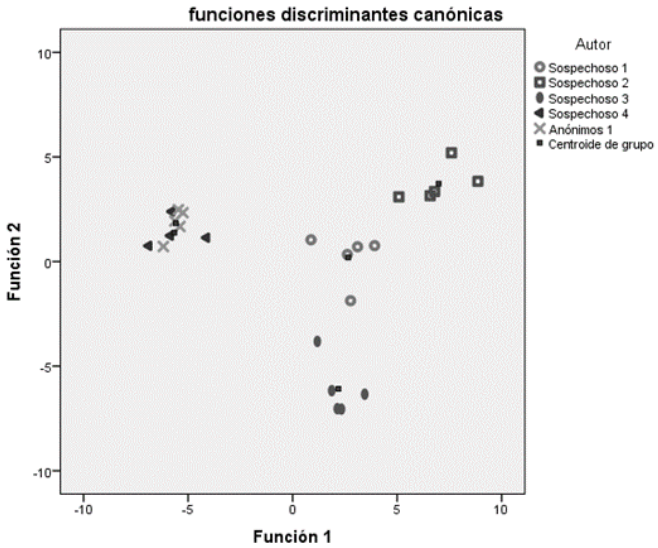


Figura 13: Resultados análisis discriminante de secuencias de etiquetas morfosintácticas

De este modo, se puede concluir que actualmente existen varias herramientas informáticas que pueden agilizar el análisis de algunas categorías lingüísticas utilizadas en la comparación forense de textos escritos. Sin embargo, los conocimientos y la experiencia del perito lingüista siguen siendo claves e imprescindibles para aplicar dichos recursos de la manera más eficiente posible y para llevar a cabo análisis rigurosos y fundamentados en teorías científicas validadas.

7. Reflexiones finales: el lingüista forense como profesional

Los distintos ejemplos tratados en esta comunicación ilustran algunas de las tareas que deben resolver los lingüistas forenses especializados en muestras escritas. Además, se ha defendido la importancia de la formación de base de dichos especialistas, así como la necesidad de que la formación sea continua y les permita conocer en profundidad los avances teóricos, metodológicos y prácticos no solo en la disciplina de la lingüística forense, sino también en distintas disciplinas y áreas del conocimiento relacionadas y relevantes para el ejercicio de su profesión como el análisis del discurso, la lingüística computacional, la docencia de idiomas, etc.

Como se ha visto, las consecuencias (legales o de otro tipo) vinculadas a las consultas que se hacen a los lingüistas forenses pueden ser especialmente graves y variar en gran medida dependiendo, entre otros factores, de la tipología del caso. Es por ello que un lingüista forense debe ser muy consciente de las limitaciones éticas de su trabajo y ceñirse a los códigos de buenas prácticas de la Asociación Internacional de Lingüistas Forenses (IAFL, 2018) y de otras asociaciones a las que pueda pertenecer (como asociaciones de peritos judiciales). Además, debe tener presente que las conclusiones a las que llegue a través de su análisis lingüístico deben estar siempre fundamentadas en el análisis y en nada más (Queralt y Giménez, 2019: 78).

Queralt (2019: 19), en base a Butters (2009: 246), ofrece un listado que resume las características que, en su opinión, debe poseer un experto para poder ejercer de perito con profesionalidad:

1) Contar con una sólida formación de base en lingüística o disciplinas afines (traducción, interpretación, filología, etc.).
2) Poseer formación específica y extensa en lingüística forense. No basta con un curso de introducción a la lingüística forense o un curso de verano.
3) Tener experiencia en el ámbito de la lingüística forense.
4) Haber publicado y participado en conferencias y proyectos en los campos de la lingüística forense.
5) Mantenerse al corriente de los constantes avances teóricos y metodológicos de su disciplina.
6) Experiencia en el empleo de métodos científicos respaldados y aceptados por otros expertos en su área.
7) Pertenencia y participación en asociaciones profesionales y científicas de lingüística forense.
8) Experiencia como perito judicial en lingüística forense.

En su proyecto sobre la identidad del lingüista forense, Clarke y Kredens (2018) llevaron a cabo entrevistas con varios profesionales. Algunas de las preguntas que les hicieron se relacionan también con la dimensión ética de esta profesión. Uno de los temas recurrentes en las respuestas que recogieron es el de la importancia de conocer los límites que existen entre el rol de los lingüistas forenses y el que desempeñan otros profesionales. Así, muchos participantes en el estudio destacaron la necesidad de delimitar las pruebas que aportan como peritos a aquellas que sean necesarias en cada encargo y la de no olvidar que no es responsabilidad del lingüista llegar al fondo de los hechos acaecidos ni pronunciar un veredicto sobre ellos (2018: 95).

Además, otro de los límites sobre el que varios participantes mostraron un alto grado de conciencia fue el de convertirse en partidario de la causa del cliente. Entre las estrategias mencionadas en sus respuestas para evitar incurrir en esta falta se incluyen seleccionar métodos adecuados al material disponible y a la tarea en cuestión, evitar aceptar encargos que requieren conocimientos y experiencia fuera del alcance del perito y prescindir de información contextual sobre el encargo (2018: 94-5).

Recientemente, se han desarrollado otras iniciativas cuyo objetivo es el de reflexionar sobre el ejercicio de la lingüística forense, como el cuestionario en línea *Overview of current practices in forensic linguistics worldwide* sobre las prácticas de los lingüistas forenses a nivel mundial, elaborado por Laboratorio SQ-Lingüistas Forenses y distribuido en 2018 entre miembros de la IAFL. En total, participaron en este cuestionario 19 expertos de 13 países distintos. Los resultados obtenidos indican que algunos de los peligros que perciben los profesionales encuestados son el intrusismo y la mala praxis de ciertos peritos lingüistas. Los profesionales que participaron en el cuestionario propusieron las siguientes posibles consecuencias de dichos fenómenos para el conjunto de la profesión, todas ellas relacionadas con el desprestigio de la disciplina:

- Los tribunales pierden confianza en la opinión de los expertos.
- Nos arriesgamos a que todos los miembros de la disciplina seamos percibidos como vendedores de ciencia basura.
- Aparente falta de "cientificidad", aparente falta de objetividad.
- Desprestigiarnos a todos. ¡Y señalar sus defectos a los clientes sería poco profesional!
- Puede dañar la reputación de los lingüistas forenses y el peso concedido a los peritajes en lingüística forense.
- Tendrá consecuencias terribles para la disciplina porque dañará su reputación.
- Dará una imagen falsa de la disciplina, lo cual dificultará la aceptación cuando un profesional cualificado testifique.
- Conduce a resultados injustos en casos individuales y daña la reputación de la lingüística.

- Como en cualquier disciplina científica, los profesionales que no están suficientemente cualificados tienden a degradar el buen trabajo que los cualificados llevan a cabo.
- Posibles errores judiciales; posibles efectos adversos en la reputación de expertos cualificados; podría llevar a la pérdida de ganancias por parte de profesionales auténticos; en un caso extremo podría llevar a que los tribunales negaran la admisibilidad de peritajes forenses, como en EEUU.
- Menospreciar la disciplina porque es menos científica que otras.

En resumen, aunque las aplicaciones de la lingüística forense aquí tratadas no son las únicas posibles en esta disciplina en expansión, se han ilustrado las distintas tareas que pueden desarrollar los lingüistas forenses y cómo pueden beneficiarse de distintas aplicaciones informáticas disponibles. En relación con esta exposición de posibles análisis lingüísticos, se ha reflexionado sobre la dimensión ética de la profesión de lingüista forense, que, con la ayuda de avances tecnológicos, permite poner el conocimiento científico sobre el lenguaje a disposición de la sociedad y de la justicia.

REFERENCIAS

AUNIÓN, J. A. (2013): "No copiaré, no copiaré, no copiaré, no copiaré". *El País*. Consultado en https://elpais.com/sociedad/2013/09/18/actualidad/1379539445_080241.html.

BBC NEWS (2017): "Ciberataque masivo: ¿quiénes fueron los países e instituciones más afectados por el virus WannaCry?" Consultado en https://www.bbc.com/mundo/noticias-39929920.

BBC NEWS (MAYO 2011): "Germany's Guttenberg 'deliberately' plagiarized". Consultado en https://www.bbc.com/news/world-europe-13310042.

BUTTERS, RONALD R. (2009): "The forensic linguist's professional credentials". *The International Journal of Speech, Language and the Law*, 16(2), pp. 237-252.

CICRES, J. & QUERALT, S. (2019): "An n-gram based approach to the automatic classification of schoolchildren's writing". *Vigo International Journal of Applied Linguistics*, 16.

CLARKE, I. & KREDENS, K. (2018): "'I consider myself to be a service provider': Discursive identity construction of the forensic linguistic expert". *International Journal of Speech, Language & the Law*, 25(1), pp. 79-107.

COATES, J. (2004): *Women, men and language: A sociolinguistic account of gender differences in language*. Harlow: Pearson Longman.

DIAS, P. C. & BASTOS, A. S. (2014): "Plagiarism in Portugal – *secondary education teachers' perceptions". Procedia – Social and Behavioral Sciences,* 116, pp. 2598-2602.

ENGLISH PROFILE (2015): *About us*. Consultado en http://www.englishprofile.org/home/about-us.

FLASHPOINT (2017): "Linguistic Analysis of WannaCry Ransomware Messages Suggests Chinese-speaking Authors". Consultado en https://www.flashpoint-intel.com/blog/linguistic-analysis-wannacry-ransomware/.

GARAYZÁBAL, E., QUERALT, S. & REIGOSA, M. (2019): *Fundamentos de la Lingüística Forense*. Colección Lingüística. Editorial Síntesis. Madrid.

GARRIDO, M. (2018): "Shakira y su "Waka Waka": fútbol, amor y plagio". *Culto*. Consultado en http://culto.latercera.com/2018/06/14/waka-waka-shakira-futbol-amor-plagio/.

GIBBONS, J. & TURELL, M. T. (2008): *Dimensions of Forensic Linguistics*. Amsterdam y Philadelphia: John Benjamins.

HARRISON, J. & BARKER, F. (2015): *English Profile Studies 5. English Profile in Practice*. Cambridge: Cambridge University Press.

IAFL (2018): *International Association of Forensic Linguists. Code of practice*. Consultado en https://www.iafl.org/wp-content/uploads/2018/07/IAFL_Code_of_Practice_1-1.pdf.

KNIGHT, B. (2015): *Applications of English Profile.* En Harrison, Julia y Barker, Fiona. *English Profile Studies 5. English Profile in Practice.* Cambridge: Cambridge University Press.

LAVERY, U. A. (1921): "The language of the law". *American Bar Association Journal*, 7(6), 277-283. Recuperado de http://www.jstor.org/stable/25700865.

MEANINGCLOUD (s.f.): *Analizador morfosintáctico*. Consultado en http://www.mystilus.com/Analizador_morfosintactico.

MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE (MECD) (2002): *Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación.* Recuperado de https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf.

MONTERO, T. (2016): "Las universidades, en regulación del plagio, van terriblemente retrasadas". *La Voz de Galicia*. Consultado en https://www.lavozdegalicia.es/noticia/educacion/2016/03/09/universidades-regulacion-plagio-van-terriblemente-retrasadas/0003_201603G9P10992.htm.

MONTOLÍO, E. (2012): "La modernització del discurs jurídic espanyol impulsada pel Ministeri de Justícia. Presentació i principals aportacions de l'Informe sobre el lenguaje escrito". *Revista de Llengua i Dret*, 57, pp. 95-121.

MORENTE, L. M. (2017): "Las empresas afectadas por el ciberataque". *Expansión*. Consultado en http://www.expansion.com/economia-digital/2017/05/13/59171d0022601dab628b4597.html.

MEANINGCLOUD (s.f.): *Analizador morfosintáctico*. Consultado en http://www.mystilus.com/Analizador_morfosintactico.

POBLETE, C., ARENAS, L., CÓRDOVA, A., GONZÁLEZ, E. & TAPIA, D. (2018): *Estrategias en comprensión del discurso escrito en contextos jurídicos*. Valparaíso: Ediciones Universitarias de Valparaíso.

QUERALT, S. & TURELL, M. T. (2013): "A semi-automatic authorship attribution technique applied to real forensic cases involving Judgments in Spanish". En R. Sousa-Silva *et al.* (eds.). *Bridging the Gap(s) between Language and the Law: Proceedings of the 3rd European Conference of the International Association of Forensic Linguists*. Porto: Faculdade de Letras da Universidade do Porto, pp. 10-18.

QUERALT, S. (2014): "Acerca de la prueba lingüística en atribución de autoría hoy". *Revista de Llengua i Dret,* 62, pp. 35-48.

QUERALT, S., MARQUINA, M. & GIMÉNEZ, R. (2018): "Evidencias lingüísticas del plagio en el periodismo español". *Estudios sobre el Mensaje Periodístico* 24(2), pp. 1559-1578.

QUERALT, S. (2020): *Atrapados por la lengua*. Editorial Larousse. Barcelona.

QUERALT, S. & GIMÉNEZ, R. (2019): *Soy lingüista, lingüista forense*. Editorial Pie de Página, colección Tinta Roja. Madrid.

QUERALT, S. (2019): *Decálogo para solicitar una pericial lingüística*. Editorial Pie de Página, colección Tinta Roja. Madrid.

QUERALT, S. & GIMÉNEZ, R. (2018): "La imitación como contraargumento en peritajes de atribución de autoría: estudio de un caso". *Estudios de Lingüística Aplicada*, 68.

REAL ACADEMIA ESPAÑOLA (2009): *Nueva gramática de la lengua española*. Madrid: Espasa.

REAL ACADEMIA ESPAÑOLA (2017): *Diccionario de la lengua española*. Consultado en http://www.rae.es/rae.html.

REYES, G. (2018): *Palabras en contexto. Pragmática y otras teorías del significado*. Madrid: Arco Libros.

SILVA-CORVALÁN, C. (2001): *Sociolingüística y pragmática del español*. Washington: Georgetown University Press.

SORIANO, J. (2018): "Primeras condenas a Tenorio por despidos improcedentes". *Hoy*. Recuperado de https://www.hoy.es/extremadura/primeras-condenas-tenorio-20180516003542-ntvo.html.

STYGAL, G. (2010): "Legal Writing: Complexity". En Coulthard, Malcolm y Johnson, Alison (eds.). *The Routledge Handbook of Forensic Linguistics*. Abingdon y Nueva York: Routledge.

TIERSMA, P. M. (1999): *Legal language*. Chicago y Londres: University of Chicago Press.

TURELL, M. T. (2010): "The use of textual, grammatical and sociolinguistic evidence in forensic text comparison". *International Journal of Speech, Language & the Law*, 17(2), 211-250.

TURELL, M. T. (2011): "La tasca del lingüista detectiu en casos de detecció de plagi i determinació d'autoria de textos escrits". *Llengua, Societat i Comunicació*, 9, pp. 69-85. Consultado en http://www.ub.edu/cusc/revista/lsc/hemeroteca/numero9/articles/9_MTTurell_ling-forense_LSC-DEF.pdf.

TURNITIN (2016): *Melania Trump Trumped by Plagiarism?* Consultado en https://www.turnitin.com/blog/melania-trump-trumped-by-plagiarism.

WEBLINGUA (s.f.): *Text Inspector*. Consultado en https://textinspector.com.