

# La verificación de autoría en la escritura conjunta de demandas

UNA PERSPECTIVA INTEGRADORA

## Authorship verification in the joint drafting of lawsuits

AN INTEGRATED APPROACH

**VICTORIA GUILLÉN-NIETO**

*Universidad de Alicante*

victoria.guillen@ua.es

<https://orcid.org/0000-0002-0742-1890>

**Resumen:** Este artículo aborda la tarea de verificación de autoría en la escritura conjunta de demandas. Las preguntas de investigación a las que se pretende dar respuesta son: ¿Qué problemas dificultan la tarea de verificación de autoría en la escritura conjunta de demandas? ¿Cómo se puede discriminar la escritura individual del sujeto cuya autoría se ha de verificar? ¿Qué método permite la validación objetiva de los resultados? El artículo se divide en dos partes. En la primera parte, se abordan los problemas que dificultan la tarea de verificación de autoría en la escritura conjunta de demandas. En la segunda parte, se propone un enfoque integrador y se aplica a un caso real juzgado en un tribunal superior de justicia español.

**Palabras clave:** Lingüística forense, identificación de autor, verificación de autor, estilística forense, estilometría forense, análisis estadístico, peritaje lingüístico, demanda.

**Abstract:** This article tackles the task of authorship verification in the joint drafting of lawsuits. The article attempts to answer the following questions: What problems hinder the task of authorship verification in the joint drafting of lawsuits? How can one discriminate the individual writing of the subject whose authorship is to be verified? What method allows for objective validation of the results? The article is divided into two parts. The first part addresses the problems hindering the task of authorship verification in the joint drafting of lawsuits. The second part proposes an integrated approach and applies it to a real case tried in a Spanish high court of justice.

**Keywords:** Forensic linguistics, authorship identification, authorship verification, forensic stylistics, forensic stylometry, statistical analysis, language expertise, lawsuit.

## I. Introducción

La lingüística forense es una rama de la lingüística aplicada cuyo objeto de estudio es el lenguaje como evidencia (Guillén-Nieto y Stein 2022: 1-33). Fue el lingüista sueco Svartvik (1968)<sup>1</sup> quien dio nombre a la disciplina cuando, a finales de los años 60 del siglo xx, actuó como perito en un tribunal de justicia británico, con motivo de la reapertura del caso Timothy John Evans. El trabajo pionero en lingüística forense de lingüistas como Svartvik (1968) y Coulthard<sup>2</sup> (2005) ha permitido que la disciplina se haya unido al conjunto de las ciencias forenses. Hoy en día, en los congresos organizados por la *Asociación Americana de Ciencias Forenses*, la lingüística forense se incluye, junto con la psicología y la psiquiatría, en el área de ciencias del comportamiento humano.

La lingüística forense tiene un reconocimiento más extendido en el derecho anglosajón que en el derecho continental porque en el primero, el sistema acusatorio promueve la aportación de periciales, incluidas las lingüísticas, a instancia de parte (cf. Hammel, 2022: 64-69). Además, se ha de tener en cuenta que la lingüística forense ha evolucionado de forma desigual en las distintas jurisdicciones del derecho continental. España, por ejemplo, es uno de los estados europeos donde la pericial lingüística se puede pedir de oficio o a instancia de parte (cf. Fernández-López, 2022). Los casos en los que se suele intervenir están asociados a la identificación de autor, la identificación de locutor, la detección de plagio, la credibilidad de testimonios, la violencia de género, el acoso laboral, entre otras posibilidades.

---

<sup>1</sup> En *The Evans Statements* (Part II), Svartvik describió el método que diseñó para verificar la autoría de cuatro declaraciones hechas por Timothy John Evans a la policía. El lingüista sueco se centró en el análisis del nivel sintáctico de la lengua, con especial atención a los tipos de oraciones presentes en las cuatro declaraciones de Evans. El análisis cualitativo se completó con un análisis estadístico (el test Chi-cuadrado de Pearson). En este caso, Svartvik demostró que los tipos de oraciones presentes en los párrafos de las declaraciones donde Evans se inculpaba del asesinato de su mujer e hija no eran propios del idiolecto de un joven analfabeto y con bajo coeficiente intelectual sino del sociolecto del agente de policía que le tomó declaración.

<sup>2</sup> En la década de los años 90 del siglo xx, Coulthard actuó como perito en la reapertura de otro caso muy similar al de Timothy John Evans en el Reino Unido: el caso Derek Bentley. Haciendo uso de herramientas como el análisis del discurso, la lingüística del corpus y la estadística descriptiva, Coulthard demostró que las declaraciones de Bentley contenían marcadores sintácticos característicos de la redacción de informes policiales, como, por ejemplo, la posición del adverbio *then* («entonces») entre el sujeto y el verbo en vez de en posición inicial de oración. Por ejemplo: *He then said* («Él entonces dijo») en vez de *Then he said* («Entonces él dijo»).

El área de identificación de autoría sigue siendo una de las más conocidas del peritaje lingüístico que, incluso, a veces se llega a confundir con la propia disciplina de la lingüística forense. Debido a una tendencia social decreciente a escribir a mano, los tribunales van requiriendo cada vez menos de la asistencia profesional de peritos calígrafos. Por experiencia como perito, lo habitual es que los textos cuya autoría se cuestiona estén escritos en soporte electrónico, ya sea email, WhatsApp, sms, o redes sociales.

Dentro del área de identificación de autor, son varias las tareas que se solicitan del perito lingüista (Fobbe, 2022: 185-186; van Halteren, 2022: 220):

- a) Atribución. El perito ha de calcular el grado de probabilidad de que una muestra de textos dubitados haya sido escrita por uno de los posibles candidatos. En esta tarea se requiere la comparación del conjunto de textos dubitados con cada uno de los conjuntos de textos indubitados de los posibles candidatos (Chaski, 1997: 15-22; 2001: 1-65; 2005: 1-13; 2006: 505-522; Grant, 2007: 1-25; 2008: 215-229; Stamatatos, 2009: 538-556; Kredens y Coulthard, 2012: 504-516; Demir, 2017: 8-12; Ehrhardt, 2018: 169-200; Zaśko-Zielińska, 2022: 373).
- b) Identificación. El perito ha de identificar al autor de una muestra de textos dubitados en una población. Los investigadores suelen hacer uso de información de tipo extralingüístico para reducir el número de sospechosos de los que se han de obtener muestras de escritura para compararlas con los textos cuya autoría se desconoce (De Jong-Lendle, 2022: 285-288).
- c) Verificación. El perito ha de verificar si un sujeto ha podido escribir, o no, una muestra de textos dubitados (Halvani, Winter y Pflug, 2016: 33-43).
- d) Perfil lingüístico: El perito ha de analizar la forma de expresión de un sujeto para intentar identificar su edad, lengua materna, dialecto, nivel de estudios, profesión u otras características que faciliten su identificación.
- e) Ofuscación: El perito ha de determinar si un sujeto ha podido ocultar rasgos de su idiolecto para evitar ser identificado. Se ha de comparar la muestra de textos dubitados con una muestra de textos indubitados del sospechoso.
- f) Imitación: El perito ha de determinar si un sujeto ha podido imitar la forma de expresión de otro sujeto para, por ejemplo,

suplantarle o incriminarle. En esta tarea, se ha de comparar la muestra de textos dubitados con una muestra de escritura del sospechoso y otra del sujeto que se cree ha sido imitado.

Lejos de ser independientes, las tareas de identificación referidas se encuentran, unas más que otras, relacionadas entre sí. Por ejemplo, la atribución puede considerarse como una secuencia de problemas de verificación con respecto a cada uno de los posibles candidatos, que se resuelve cuando la autoría de los textos dubitados se atribuye a uno de los sujetos.

Este artículo se centra en la tarea de verificación de autoría en escritura conjunta. Y además lo hace en un género profesional específico: la demanda. Las preguntas de investigación a las que el artículo pretende dar respuesta son: ¿Qué problemas dificultan la tarea de verificación de autoría en la escritura conjunta de demandas? ¿Cómo se puede discriminar la escritura individual del sujeto cuya autoría se ha de verificar? ¿Qué método permite la validación objetiva de los resultados? En la primera parte, se abordan los problemas que dificultan la tarea de verificación de autoría en la escritura conjunta de demandas. Empezaré abordando el frágil concepto de autoría en la escritura conjunta de demandas. Luego explicaré los problemas que dificultan la tarea de verificación de autoría en la escritura conjunta de una demanda. La primera parte termina haciendo una referencia somera a los principales enfoques y métodos en la identificación de autor. En la segunda parte, propondré un enfoque integrador y lo aplicaré a un caso real de verificación de autoría en escritura conjunta de demandas, que fue juzgado en un tribunal superior de justicia español.<sup>3</sup>

## **2. La autoría en escritura conjunta**

Love (2002) definió la autoría como un conjunto de actividades relacionadas que a veces son hechas por una sola persona, pero que en la mayoría de las ocasiones son llevadas a cabo en colaboración o por varias personas sucesivamente. Love diferenció cuatro funciones de autoría: (a) precursora, (b) ejecutiva, (c) declarativa y (d) revisora. Como su nombre indica, el autor precursor es el artífice de una obra anterior que se incorpora a una obra nueva

---

<sup>3</sup> Por la Ley de Protección de Datos, se han omitido todas las referencias particulares al caso.

(Guillén-Nieto, 2020: 117-118). El autor ejecutivo es el artífice del texto, es decir, la persona que escribe el texto en la página o lo compone en el ordenador. El autor declarativo es el sujeto que firma y, al hacerlo, asume públicamente la responsabilidad de la obra por la que, además, percibe derechos morales. El revisor sugiere modificaciones para mejorar la composición del texto y supervisa el trabajo en curso del autor, pudiendo con ello modificar rasgos de la escritura individual de aquél. A partir de la definición de Love, queda claro que en la producción de un texto pueden intervenir diferentes sujetos, cada uno de ellos desempeñando una función de autoría distinta. La tarea de verificación de autoría se complica todavía más en los casos de autoría conjunta, es decir, cuando dos o más sujetos intervienen en la redacción de un texto. En tal caso, será esencial identificar rasgos de escritura que permitan discriminar a un autor de otro, con el fin de constatar si dichas diferencias de escritura se manifiestan, o no, cuando se comparan los textos indubitados de los autores con la muestra de textos dubitados.

Sin embargo, la identificación de rasgos de escritura que sirvan para diferenciar a un autor de otro no es tarea fácil en una demanda por varias razones. En la redacción de una demanda, se suelen emplear plantillas, copiar y pegar texto de un documento a otro en el mismo procedimiento civil o penal, y son frecuentes las citas a las leyes y jurisprudencia aplicables al caso, declaraciones tomadas por los agentes de policía, testimonios, informes periciales, entre otras posibilidades. Todos estos factores dificultan sobre manera la tarea de verificación de autoría porque encubren los rasgos de escritura individual del autor dubitado.

### **3. La demanda como género jurídico**

Además de las dificultades aludidas, se ha de tener en cuenta que la demanda, el acto procesal que da comienzo a un procedimiento judicial, es un género o tipo de actividad jurídica estandarizado. Lo cual significa que la demanda se caracteriza por unos rasgos genéricos y, por tanto, comunes a todos los escritos de demanda que entorpecen la tarea de verificación de autoría. Entre los rasgos genéricos de la demanda, se encuentran los siguientes (Alcaraz Varó y Hughes, 2002: 31-78):

- a) El objetivo es exigir la reconsideración de una decisión que se considera injusta y que no satisface los intereses del demandante.
- b) Tiene una superestructura preestablecida que permite al emisor y receptor del texto reconocer de qué tipo de texto se trata y cuál es su función como instrumento de comunicación profesional. La superestructura consta de los siguientes movimientos retóricos: el Encabezamiento, los Hechos, los Fundamentos de Derecho, y la Petición.
- c) El registro es formal y el tono o forma de expresión es solemne.
- d) El estilo es expositivo y argumentativo.
- e) El léxico es de tipo jurídico: se usan formalismos jurídicos, latinismos, vocabulario semitécnico, metáforas léxicas y colocaciones léxicas (Alcaraz Varó y Hughes, 2002: 31-78).
- f) La sintaxis es compleja; abundan los grupos nominales largos, las proposiciones subordinadas adjetivas o de gerundio y participio, las proposiciones de relativo, especificativas y explicativas, las formas verbales arcaicas, y las oraciones extremadamente largas debido al uso generalizado de la hipotaxis (oraciones complejas) y parataxis (oraciones compuestas).

#### 4. La tarea de verificación de autoría: enfoques y métodos

La verificación de autoría parte de una premisa lingüística: cada usuario de la lengua tiene preferencias y hábitos individuales que determinan el uso particular que hacen de la misma. A este tipo de variación individual dentro de una lengua se le denomina *idiolecto*. Los usuarios de una lengua difieren tanto en sus repertorios lingüísticos (el vocabulario y los patrones lingüísticos que conocen) como en las elecciones particulares que hacen de dichos repertorios lingüísticos cuando hablan y escriben. Sin embargo, conviene precisar que, en un caso, la muestra de textos que se tiene de un sujeto representa tan solo una pequeña parte de su idiolecto, por lo que resulta imposible acceder a todos los rasgos de escritura que facilitarían su identificación o verificación como autor. Es más, como apuntan Ainsworth y Juola (2019: 1167), no se puede demostrar empíricamente que el idiolecto de un individuo sea único. En la mayoría de los casos, el idiolecto se reduce a desviaciones inconscientes de la norma o

a una tendencia o preferencia en el uso del lenguaje por unos elementos lingüísticos sobre otros.

La tarea de verificación de autor se puede abordar desde varios enfoques y con distintos métodos a los que me referiré a continuación, aunque, por limitaciones de espacio, tan solo lo haga de forma somera.

#### 4.1 La estilística forense

En estilística forense se aplica el conocimiento lingüístico para identificar los rasgos de la escritura de un sujeto que puedan considerarse marcadores de su estilo individual, con el fin de atribuirle, o no, la autoría de un texto dubitado (Svartvik, 1968; Coulthard, 2005: 249-274; Kniffka, 2000: 179-198; Fobbe, 2020; 2022).

Los marcadores de estilo se buscan en diversos niveles (grafema, palabra, oración y texto). La selección de los marcadores estilísticos se debe hacer atendiendo a criterios objetivos para evitar lo que los expertos llaman en inglés *flower picking* («coger una flor aquí y otra allá»). En un registro formal, esta expresión coloquial equivale a decir que la selección de marcadores estilísticos se ha hecho sin base científica. Los marcadores estilísticos que pueden ser candidatos a formar parte del idiolecto de un sujeto han de ser idiosincrásicos.<sup>4</sup> Para demostrar con datos objetivables que un marcador estilístico es idiosincrásico, éste ha de ser prominente. La prominencia se puede demostrar consultando diccionarios, bases de datos y corpus lingüísticos. La experiencia demuestra que, en la mayoría de los casos, los marcado-

---

<sup>4</sup> Sobre rasgos idiosincrásicos en la escritura de un autor, léase *The UNABOMBER Manifesto. Industrial Society and Its Future* de Theodore Kaczynski, y la declaración jurada de Terry D. Turchie (*Affidavit of Assistant Special Agent in Charge*. U.S. District of Montana). En este documento se recoge el análisis lingüístico que Fitzgerald, agente del FBI e investigador forense, hizo en el caso UNABOMBER. Fitzgerald comparó el manifiesto y las cartas de amenaza cuya autoría se desconocía con una amplia muestra de cartas del principal sospechoso: Theodore Kaczynski. Fitzgerald analizó frecuencias de palabras y variantes ortográficas con un programa informático. El investigador pudo identificar a Kaczynski como autor del manifiesto y cartas de amenaza por su reconocible idiolecto. Kaczynski empleaba expresiones inusuales en inglés como *cool-headed logicians* («lógicos con la mente fría»); expresiones arcaicas, típicas de la década de los años 50, como *negro* («negro») para referirse a un ciudadano Afro-Americano, y también se desviaba de la norma. Tomemos como ejemplo el dicho: *You can't have your cake and eat it too!* («¡No se puede tener el pastel y comerlo también!»), que Kaczynski escribía como: *We can't eat our cake and have it too!* («¡No se puede comer nuestro pastel y tenerlo también!»).

res estilísticos que sirven para la verificación de autoría son el resultado de desviaciones de la lengua estándar o de una marcada tendencia o preferencia por unos elementos lingüísticos sobre otros.

Por ejemplo, a nivel grafémico, un patrón específico de puntuación, espaciado u ortografía pueden ser marcadores estilísticos. A nivel léxico, se pueden considerar marcadores estilísticos una preferencia en el uso del lenguaje por determinadas palabras sobre otras, la presencia de expresiones propias de una variedad dialectal (variedad regional dentro de una lengua) en otra, o la presencia de rasgos de un argot específico (variedad social dentro de una lengua). A nivel oracional, los marcadores estilísticos pueden incluir una preferencia por oraciones cortas sobre oraciones largas, o una marcada tendencia por el uso de unas estructuras gramaticales como, por ejemplo, la hipotaxis o la parataxis.

En cualquiera de los niveles (grafema, palabra, oración, párrafo), los marcadores estilísticos pueden analizarse de forma cualitativa (su originalidad) y evaluarse de forma cuantitativa, comparando sus frecuencias entre los textos dubitados y los indubitados. La mayor de las críticas que ha recibido la estilística forense como método es su dificultad para ofrecer una validación objetiva de los resultados (Ainsworth y Juola, 2019: 1168-1171). Por esta razón, siempre que sea posible, es aconsejable que los resultados del análisis lingüístico cualitativo se complementen con datos objetivos.

#### 4.2 La estilometría forense

La estilometría forense es la «cuantificación del estilo por medio de la búsqueda de rasgos textuales que sean medibles y representativos del estilo de un autor» (López-Escobedo, Sierra y Solórzano, 2019: 69). Por ejemplo, si se demuestra que la frecuencia de uso de determinados signos de puntuación es una característica distintiva del estilo de un autor, entonces se dice que los signos de puntuación es un marcador estilométrico. Dentro de la categoría estilométrica de signos de puntuación, se encuentran rasgos específicos como el punto, la coma, el punto y coma, entre otros.

A pesar de que se han identificado e investigado cientos de marcadores estilométricos (Holmes, 1998; Houvardas y Stamatatos, 2006: 41-83; Bergsma, Post y Yarowsky, 2012: 327-337; Brennan, Afroz y Greestadt, 2012: 12; Daelemans, 2013: 451-462), no se ha encontrado un conjunto que funcione para todas las situaciones comunicativas. De acuerdo con Stamatatos (2009:

540), los marcadores estilométricos se pueden clasificar atendiendo a varios criterios: a) caracteres, b) léxicos, c) semánticos, d) sintácticos y e) específicos a la aplicación. Algunos de los más representativos en cada nivel se exponen en la **Tabla 1**.

Criterios	Rasgos estilométricos
Caracteres	-Secuencias de caracteres de n-gramas
Léxicos	-Tokens: secuencias de palabras de n-gramas, promedio de longitud de palabra, promedio de longitud de oración, promedio de longitud de párrafo -Riqueza de vocabulario: ratio tipo/token, palabras diferentes, hápax legomena, hápax dislegomena -Frecuencias de palabras -Errores
Semánticos	-Sinónimos e hiperónimos -Dependencias semánticas -Características funcionales de las palabras en el discurso
Sintácticos	-Estructura de frases -Estructura de cláusulas -Estructura de oraciones -Estructura de párrafos
Específicos a la aplicación	-Estructuras específicas del contenido -Estructuras específicas de la expresión -Funciones -Contenido específico -Idioma específico

**Tabla 1.** Clasificación de elementos estilométricos (adaptada de Stamatatos, 2009: 540)

Ainsworth y Juola (2019: 1171) argumentan que la estilometría forense resuelve el problema de validación objetiva requerido en los tribunales porque ofrece un método de análisis que puede ser repetido, replicado y comprobado.

### 4.3 La lingüística forense computacional

La lingüística forense computacional se centra en la extracción automática de rasgos lingüísticos que puedan ser cuantificados. Como explica van Halteren (2022), se pueden emplear bien listados de rasgos ya existentes o bien hacer uso de la observación e intuición para decidir qué rasgos se han de extraer en cada caso concreto. A continuación, se calculan las frecuencias absolutas y relativas, y la variación de frecuencia básica de los rasgos lingüísticos extraídos. También se pueden aplicar medidas de riqueza léxica como ratio tipo/token, palabras diferentes, hápax legomena, hápax dislegomena, entre otras. Después se extraen los vectores de rasgos lingüísticos de los textos, y se comparan empleando aprendizaje automático, árboles de decisión y pruebas estadísticas como, por ejemplo, el *análisis discriminante lineal* (van Halteren, 2022: 236-237). Hoy en día, cuando el número de rasgos lingüísticos que se ha de analizar es muy grande (*Big data*), se emplea el llamado *aprendizaje profundo* (véase van Halteren, 2022: 238-239). Para poder implementar esta técnica avanzada de inteligencia artificial con fiabilidad, es un requisito indispensable disponer de datos masivos, algo que, desafortunadamente, no suele ocurrir en los casos forenses reales (Fobbe, 2022: 187-189).

### 5. Una aproximación integradora a un estudio de caso sobre verificación de autoría en escritura conjunta de demandas

La perspectiva que se adoptó en este caso de verificación de autoría integró el análisis de aspectos extralingüísticos, lingüísticos, estilométricos y estadísticos. El análisis lingüístico cualitativo se complementó con el análisis cuantitativo para aportar datos objetivables. El método aplicado se puede repetir, puede ser replicado por otro investigador, y aporta una validación objetiva.

Es recomendable empezar el estudio de caso definiendo el *acontecimiento de habla* en el que se encuadra (Guillén-Nieto, 2020: 115). El acontecimiento de habla es un modelo o esquema mental que comparten los usuarios de la lengua y facilita una comunicación adecuada. El acontecimiento de habla se define a partir de una serie de variables contextuales que van Dijk (2008: 2) resume en cinco: (a) el contexto, (b) los participantes, (c) las acciones, (d) los objetivos, (e) el conocimiento. El contexto del caso que nos ocupa era profesional. Los participantes eran un abogado argentino y un abogado español. El

primero tenía un contrato de colaboración de servicios jurídicos en un bufete cuyo director era el segundo. En consecuencia, entre ambos abogados existía una relación jerárquica. Las acciones llevadas a cabo fueron, entre otras, la colaboración en prestación de servicios jurídicos, la preparación de una serie de casos, la búsqueda de leyes y jurisprudencia aplicables a los casos, y la redacción conjunta de demandas. Los dos abogados trabajaron juntos en numerosos casos, por lo que en las demandas figuraban las firmas de ambos como autores (autoría declarativa). Según la Ley de Propiedad Intelectual española, la firma de un documento declara la autoría del mismo:

Artículo 6. Presunción de autoría, obras anónimas o seudónimas. 1.  
Se presumirá autor, salvo prueba en contrario, a quien aparezca como tal en la obra, mediante su nombre, firma o signo que lo identifique.  
(Ley de Propiedad Intelectual 1/1996 del 12 de abril de 1996).

El conflicto surgió cuando el director del bufete se negó a conceder los honorarios correspondientes al abogado, alegando que éste no había colaborado en la redacción de las demandas (autoría ejecutiva). En consecuencia, el abogado se vio forzado a interponer demanda contra el director del bufete por incumplimiento de contrato e infracción de daños morales. Para admitir la demanda, el tribunal puso como condición que se verificara la autoría del demandante en los textos jurídicos cuya autoría se discutía. Conviene precisar que la autoría del abogado demandado (el director del bufete) no se puso en duda. El tribunal designó de oficio a un perito lingüista para verificar la autoría del demandante.

La muestra que se entregó al perito lingüista consistió en treinta y seis demandas (182 496 palabras), divididas en tres conjuntos de textos: (a) doce demandas dubitadas firmadas conjuntamente por ambos abogados, demandante y demandado (72 418 palabras), (b) doce demandas indubitadas del demandante (87 743 palabras), y (c) doce demandas indubitadas del demandado (22 308 palabras). Aunque la similitud en el número de palabras entre las demandas dubitadas y las indubitadas del demandante podía considerarse un indicio de autoría, este hecho lingüístico no era evidencia suficiente para su verificación en un tribunal de justicia.

A pesar de que la verificación de autoría en escritura conjunta es siempre una tarea ardua, el caso presentaba algunas ventajas para la investigación: (a) se trataba de un caso cerrado porque solamente se tenía que verificar la autoría de uno de los abogados firmantes; (b) el tamaño de la muestra era grande (182 469 palabras); (c) la posibilidad de analizar demandas indubitadas de cada abo-

gado; y (d) la uniformidad de la muestra, es decir, todos los textos proporcionados para el análisis pertenecían al mismo género y tipo de texto jurídico, esto es, la demanda.

Sin embargo, el caso también presentaba algunas desventajas. Por ejemplo, el hecho de que ambos abogados trabajaran en el mismo bufete significaba que podían usar la misma plantilla para redactar las demandas, podían redactarlas conjuntamente, e incluso copiar y pegar texto de un documento a otro dentro del mismo procedimiento. Además, la redacción conjunta de las demandas podía estar revisada por el abogado de mayor estatus profesional. En la investigación, se plantearon dos hipótesis:

- a) Hipótesis nula: Se predijo que no existía una relación de dependencia entre la escritura del demandante y la redacción de las demandas dubitadas.
- b) Hipótesis alternativa: Se predijo que sí existía una relación de dependencia entre la escritura del demandante y la redacción de las demandas dubitadas.

### 5.1 Procedimiento

En primer lugar, se redujo la muestra de textos original de 182 496 palabras a 15 000 palabras, con la finalidad de eliminar, en la medida de lo posible, los formalismos jurídicos y las citas que pudieran entorpecer la tarea de verificación de autoría. La muestra resultante era más homogénea y uniforme porque todos los textos tenían una extensión de 500 palabras y se expresaban en un registro expositivo. Todos los textos fueron extraídos de los «Hechos», que, presumiblemente, es la parte de la demanda con escritura más espontánea. Más específicamente, la nueva muestra consistía en tres conjuntos de 10 textos de 500 palabras cada uno. En total: (a) 5000 palabras de las demandas dubitadas, (b) 5000 palabras de las demandas indubitadas del demandante, y (c) 5000 palabras de las demandas indubitadas del demandado.

En segundo lugar, se adoptó la siguiente estrategia: a) comparar las demandas indubitadas del demandante y demandado para identificar rasgos que permitieran discriminar la escritura individual de cada uno de ellos, b) comparar las demandas indubitadas del demandante con las demandas dubitadas para comprobar si las diferencias observadas entre la escritura indi-

vidual del demandante y demandado se mantienen o no, c) en caso negativo, se ha de comprobar si dichas diferencias se hallan, o no, cuando se comparan las demandas indubitadas del demandado con las demandas dubitadas, d) en el caso de que sí se hallen diferencias entre la escritura individual del demandado y la escritura individual del autor de las demandas dubitadas, el marcador, estilístico o estilométrico, divergente se podrá considerar un indicador de autoría del demandante.

En tercer lugar, se decidió que el análisis se centrara en los niveles léxico y sintáctico. En cuanto al nivel léxico, la idea no era encontrar preferencias por unos términos jurídicos sobre otros, ya que éstos forman parte del sociolecto de los profesionales del derecho. Se pensó que, dado el diferente origen geográfico del demandante (argentino) y demandado (español), la búsqueda, en unas demandas supuestamente escritas en español peninsular, de algún rasgo dialectal del español de América que se hubiera pasado por alto en la revisión de los textos podría ser más fructífera para la tarea de verificación de autoría. El objetivo era diferenciar la escritura de demandante y demandado por la presencia de rasgos del español de América, para luego verificar si dichos rasgos se encontraban también en las demandas dubitadas.

En cuanto al nivel sintáctico, tomando como fuente de inspiración el modelo de análisis sintáctico de Svartvik (1968), se analizó la complejidad oracional en los tres conjuntos de textos de la muestra. Se pensó que, como cada sujeto *empaqueta* sus ideas y pensamientos eligiendo diferentes estructuras sintácticas, era posible encontrar en las escrituras del demandante y demandado diferencias en la tendencia o preferencia por unos tipos de oraciones sobre otros. Además, por su base cognitiva, el nivel sintáctico es más difícil de modificar por un revisor que los niveles grafémico y léxico. Primero, se clasificaron las oraciones en cuatro grandes categorías: (1) Tipo A (oración simple), (2) Tipo B (oración compuesta), (3) Tipo C (oración compleja) y (4) Tipo D (compuesta-compleja). A continuación, se calcularon sus frecuencias absolutas.

En quinto lugar, se hizo un análisis sintáctico más detallado. Esta vez, los tipos de oraciones se dividieron en subtipos:

Tipo A (oración simple): Declarativa (A1), Interrogativa (A2), Imperativa (A3), Exclamativa (A4)

Tipo B (oración compuesta): Copulativa (B1), Adversativa (B2), Disyuntiva (B3), Aclarativa (B4)

Tipo C (oración compleja): Subordinada sustantiva (C1), Subordinada adjetiva o de relativo (C2), Especificativa (C21), Explicati-

va (C22); Subordinada adverbial (C3), Tiempo (C31), Lugar (C32), Modo (C33), Comparativa (C34), Subordinada causal (C35), Subordinada consecutiva (C36), Subordinada condicional (C37), Subordinada contrastiva (C38) Subordinada final (C39).

En sexto lugar, se analizaron los siguientes marcadores estilométricos: palabras diferentes, ratio tipo/token, promedio de longitud de palabra, promedio de longitud de oración y secuencias de palabras de n-gramas.

Finalmente, se hizo un análisis inter-autor para lo que se practicó un análisis de varianza (ANOVA)<sup>5</sup>. El objetivo era determinar si, al comparar cada conjunto de textos de la muestra con los demás, se observaban diferencias significativas en cuanto a las frecuencias de los tipos de oraciones, tipos de oraciones subordinadas, y variables estilométricas arriba indicadas.

## 5.2 Resultados

Por limitaciones de espacio, en esta sección solamente expondremos los resultados que fueron determinantes para la verificación de autoría del demandante.

### 5.2.1 Nivel léxico: Rasgos léxicos del español de América

A partir de una lectura extensiva de los textos dubitados, se identificó la presencia de palabras como *Defensoría del pueblo*, *puntaje* y la locución preposicional *luego de*. Se consultó, en primer lugar, el *Diccionario de la lengua española (DLE)*. Las entradas del diccionario confirmaron que se trataba de expresiones del español de América:

---

<sup>5</sup> El análisis de varianza (ANOVA) prueba la hipótesis de que las medias de dos o más grupos son iguales. Este tipo de análisis hace uso de las varianzas para determinar si las medias son diferentes. El procedimiento funciona comparando la varianza entre las medias de los grupos y la varianza dentro de los grupos como una forma de determinar si los grupos son todos parte de una población más grande o poblaciones separadas con características diferentes. Si el valor *p* es menor que el nivel de significancia, se concluye que al menos una media es diferente.

- a) *Luego de* es una locución preposicional cuyo significado es «con posterioridad en el tiempo».
- b) *Defensoría* es un sustantivo típico cuyo significado es «Ministerio o ejercicio de defensor».
- c) *Puntaje* es un sustantivo cuyo significado es «conjunto de puntos obtenidos en algún tipo de prueba».

A continuación, se buscaron dichas expresiones y sus equivalentes en español peninsular (*después de*, *Defensor del Pueblo* y *puntuación*) en el *Corpus del Español del Siglo XXI* (CORPES). La búsqueda sirvió para establecer, con datos objetivables, las diferencias entre dos variedades del español, el español americano y el español peninsular (**Tabla 2**).

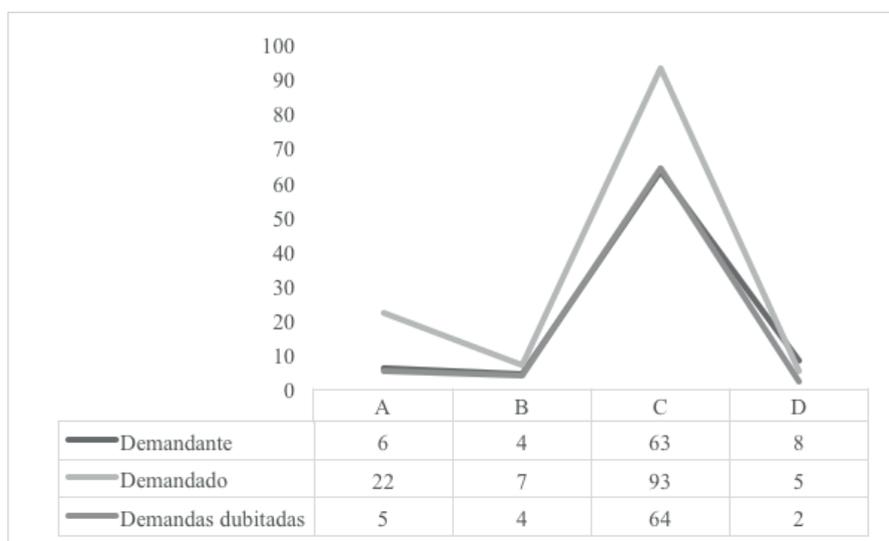
Variedad geográfica o dialectal	Distribución por país	Expresión	Frecuencia
Español de América	Argentina	<i>Luego de</i>	6026
	España	<i>Luego de</i>	721
	Argentina	<i>Defensoría</i>	69
	España	<i>Defensoría</i>	10
	Argentina	<i>Puntaje</i>	106
	España	<i>Puntaje</i>	7
Español peninsular	Argentina	<i>Después de</i>	14 445
	España	<i>Después de</i>	54 657
	Argentina	<i>Defensor del Pueblo</i>	24
	España	<i>Defensor del Pueblo</i>	356
	Argentina	<i>Puntuación</i>	63
	España	<i>Puntuación</i>	1332

**Tabla 2.** Comparación entre Argentina y España en la distribución de frecuencias absolutas de expresiones (CORPES)

Los datos expuestos en la **Tabla 2** sirvieron para demostrar la presencia de rasgos del español de América en las demandas objeto de la dispu-

ta. Es significativo que las tres expresiones también estuvieran presentes en el conjunto de demandas indubitadas del demandante (abogado argentino) pero no en las del demandado (abogado español). En consecuencia, la presencia o no presencia de rasgos dialectales del español de América en los tres conjuntos de textos sirvió, en este caso, para diferenciar la escritura de los dos abogados y, en consecuencia, se consideró un indicador válido para la verificación de autoría del demandante.

### 5.2.2 Nivel sintáctico: tipos de oraciones



**Gráfico 1.** Distribución por conjunto de textos de las frecuencias absolutas de los tipos de oraciones.

El **Gráfico 1** muestra una estrecha similitud entre los textos indubitados del demandante y las demandas dubitadas en relación con las frecuencias absolutas de los tipos A, B y C. Por el contrario, esta similitud se disipa cuando se comparan los textos indubitados del demandado y las demandas. Las demandas indubitadas del demandante contienen un número considerablemente menor de oraciones complejas ( $N = 63$ ) que las del demandado ( $N = 93$ ). Con el fin de confirmar si las diferencias observadas eran, o no, estadísticamente significativas, se realizó un ANOVA (IBM SPSS v. 26). Este estadístico arrojó los siguientes resultados:

- a) Se hallaron diferencias altamente significativas entre las demandas indubitadas del demandante y las del demandado en lo referente a la oración compleja (tipo C).

$$F (M = 6,3; DS = 1,1, M = 9,3; DS = 2,2) = 15,5, p < 0,00$$

- b) No se hallaron diferencias significativas entre las demandas indubitadas del demandante y las demandas dubitadas en ningún tipo de oración (A, B, C y D). En concreto, los resultados con respecto a la oración compleja (tipo C) fueron:

$$F (M = 6,3; DS = 2,1, M = 6,4; DS = 2,5) = 0,00, p = 0,9$$

- c) Se encontraron diferencias altamente significativas entre las demandas indubitadas del demandado y las demandas dubitadas en cuanto a la oración compleja (tipo C).

$$F (M = 9,3; DS = 1,1, M = 6,4; DS = 2,5) = 10,7, p = 0,00$$

Los resultados evidenciaron variación inter-autor en relación con el uso de la oración compleja (tipo C). Más específicamente, bajo las mismas circunstancias (un texto expositivo de 500 palabras), el demandante empleó un número considerablemente menor de oraciones tipo C que el demandado. De ello se deduce que el tipo C, además de ser un rasgo genérico de los textos jurídicos, en este caso sirvió para diferenciar la escritura individual del demandante y demandado, y, por lo tanto, pudo considerarse como otro indicador válido para la verificación de autor.

### 5.2.3 Nivel sintáctico: subtipos de oraciones subordinadas

Una vez divididos los tipos de oraciones por subtipos, se calcularon sus frecuencias absolutas en cada uno de los tres conjuntos de textos de la muestra. El resultado del cálculo arrojó diferencias visibles en el subtipo C<sub>I</sub>:

- a) En las demandas dubitadas, la frecuencia absoluta del subtipo C<sub>I</sub> (subordinada sustantiva) era  $N = 39$ .
- b) En los textos indubitados del demandante, la frecuencia absolu-

ta del subtipo C<sub>I</sub> (subordinada sustantiva) era  $N=34$ .

- c) En los textos indubitados del demandado, la frecuencia absoluta del subtipo C<sub>I</sub> (subordinada sustantiva) era  $N=80$ .

Para comprobar si las diferencias observadas eran significativas desde un punto de vista estadístico, se hizo un ANOVA (IBM SPSS v. 26), cuyos resultados fueron los siguientes:

- a) Cuando se compararon las demandas indubitadas del demandante con las del demandado, se encontraron diferencias altamente significativas en la frecuencia de uso de la oración subordinada sustantiva (subtipo C<sub>I</sub>).

$$F (M = 3,9; DS = 1,3, M = 8; DS = 2,3) = 23,3, p < 0,00$$

- b) Cuando se compararon las demandas indubitadas del demandante con las demandas dubitadas, no se encontraron diferencias significativas en la frecuencia de uso de la oración subordinada sustantiva (subtipo C<sub>I</sub>).

$$F (M = 3,9; DS = 1,3, M = 3,4; DS = 2,5) = 0,30, p = 0,58$$

- c) Cuando se compararon las demandas indubitadas del demandado con las demandas dubitadas, se encontraron diferencias altamente significativas en la frecuencia de uso de la oración subordinada sustantiva (subtipo C<sub>I</sub>).

$$F (M = 8; DS = 2,3, M = 3,4; DS = 2,5) = 18,24, p < 0,00$$

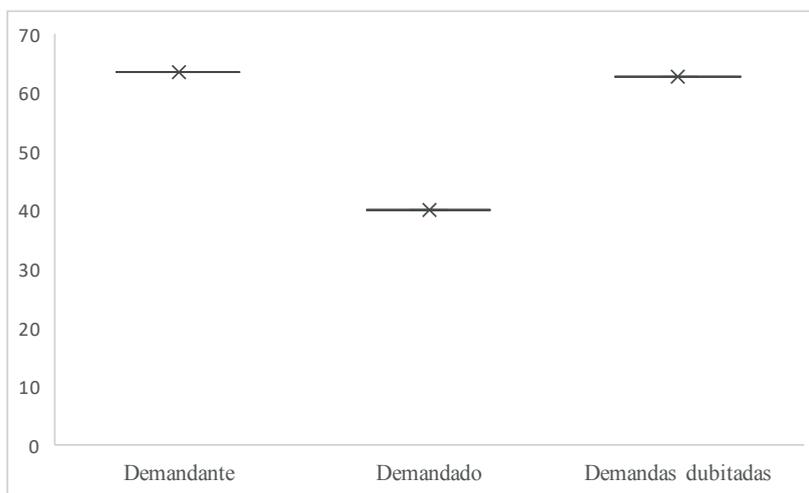
En consecuencia, el subtipo C<sub>I</sub> (subordinada sustantiva) pudo ser considerado un indicador válido para la verificación de autoría del demandante.

#### 5.2.4 Nivel sintáctico: Promedio de longitud de oración

Los marcadores estilométricos se calcularon automáticamente con la herramienta *Text Works* (Gil-Salom, Soler-Monreal, Stuart y Candela-Coderch 2004). Los resultados fueron los siguientes:

- a) Promedio de longitud de oración de las demandas dubitadas: 62,91 palabras.
- b) Promedio de longitud de oración de las demandas indubitadas del demandante: 63,52 palabras.
- c) Promedio de longitud de oración de las demandas indubitadas del demandado: 40,07 palabras.

Los resultados ponen de manifiesto que mientras el promedio de longitud de oración de los textos indubitados del demandante es visiblemente semejante al de las demandas dubitadas, este promedio se diferencia del correspondiente a las demandas indubitadas del demandado. El **Gráfico 2** muestra la distribución del promedio de longitud de oración en los tres conjuntos de textos.



**Gráfico 2.** Distribución por autores y tipos de oraciones del promedio de longitud de oración

Con el fin de determinar si las diferencias observadas eran estadísticamente significativas, se recurrió a un ANOVA (IBM SPSS v. 26), que arrojó los siguientes resultados:

- a) Cuando se compararon las demandas indubitadas del demandante con las del demandado, se hallaron diferencias altamente significativas.

$$F (M = 63,5; DS = 19,1, M = 40; DS = 9,2) = 12,1, p = 0,00$$

- b) Cuando se compararon las demandas indubitadas del demandante con las demandas dubitadas, no se hallaron diferencias significativas.

$$F (M = 63,5; DS = 19,1, M = 62,9; DS = 24,3) = 0,04, p = 0,91$$

- c) Cuando se compararon las demandas indubitadas del demandado con las demandas dubitadas, se hallaron diferencias altamente significativas.

$$F (M = 40,0; DS = 9,2, M = 62,9; DS = 24,3) = 7,7, p = 0,01$$

Además del promedio de longitud de oración, se halló otro marcador estilométrico que permitió verificar la autoría del demandante: la secuencia de palabras de 4-gramas. Los resultados del ANOVA (IBM SPSS v. 26) se describen a continuación:

- a) Cuando se compararon las demandas indubitadas del demandante con las del demandado, se hallaron diferencias altamente significativas.

$$F (M = 20,1; DS = 3,2, M = 27,9; DS = 2,7) = 33,4, p = 0,00$$

- b) Cuando se compararon los textos indubitados del demandante con las demandas dubitadas, no se hallaron diferencias significativas.

$$F (M = 20,1; DS = 3,2, M = 22,60; DS = 5) = 1,7, p = 0,20$$

- c) Cuando se compararon las demandas indubitadas del demandado con las demandas dubitadas, se hallaron diferencias altamente significativas.

$$F (M = 27,9; DS = 2,7, M = 22,6; DS = 5) = 8,5, p = 0,00$$

El promedio de longitud de oración y las secuencias de palabras de 4 gramas demostraron una variación inter-autor valiosa para discriminar la escritura de los abogados contendientes y verificar la autoría del demandante. En resumen, los marcadores, de estilo y estilométricos, que permitieron la verifica-

ción de autoría del demandante fueron: a) la presencia de rasgos del español de América; b) la semejanza en el patrón de uso de los tipos de oración A, B, C, D; c) el uso menos frecuente de la oración subordinada (subtipo C); d) el uso menos frecuente de la oración subordinada sustantiva (subtipo C1); e) el promedio de longitud de oración (las oraciones del demandante son más largas que las del demandado); y f) el uso más frecuente de la secuencia de palabras de 4-gramas.

En consecuencia, en la investigación se rechazó la hipótesis nula y aceptó la hipótesis alternativa, que predecía una relación de dependencia entre la escritura del demandante y la redacción de las demandas dubitadas.

## 6. Conclusiones

Este artículo abordó el problema de la verificación de autoría en la escritura conjunta de demandas. El artículo puso de manifiesto las dificultades a las que el perito ha de hacer frente. Unas dificultades son inherentes a la demanda como género profesional jurídico. Es decir, todas las demandas comparten unos rasgos genéricos (objetivo, superestructura, léxico especializado y sintaxis) que permiten a los usuarios de la lengua reconocer este instrumento de comunicación profesional en el ámbito jurídico. Otras dificultades se deben al uso de plantillas, a la acción de copiar y pegar texto de un documento a otro en el mismo procedimiento civil o penal, y a la intertextualidad en forma de abundantes citas a leyes, jurisprudencia, sentencias, informes forenses, entre otros documentos a los que se puede hacer referencia.

El artículo demostró que es posible verificar la escritura conjunta de demandas a partir de un estudio estratégico de los niveles que más convenientes y fructíferos puedan ser para la investigación. Es en este punto de la investigación donde además de sólidos conocimientos lingüísticos, se requiere capacidad de observación, intuición y creatividad. Se trata de seguir el rastro de un elemento para convertirlo en un hecho lingüístico que el tribunal pueda validar como evidencia. No existe una plantilla de análisis que se pueda aplicar a todos los casos de verificación de autoría por igual. Tampoco existe un programa de ordenador que pueda hacer el trabajo por el perito lingüista. Cada caso es distinto y requiere un estudio independiente. A veces, el método de análisis empleado es inadecuado o poco productivo, y se ha de rediseñar el modelo de análisis.

En el estudio de caso presentado, se tomó la decisión de reducir la amplia muestra de textos (182 496 palabras) a una más homogénea y uniforme

(15 000 palabras), extrayendo 500 palabras de la sección «Hechos» por considerarse su escritura más espontánea que la de otras secciones de la demanda. Otra decisión importante fue analizar los niveles léxico y sintáctico. En el primer caso, el conocimiento extralingüístico de la distinta procedencia geográfica del demandante (argentino) y demandado (español), hizo pensar que, si el demandante había colaborado en la redacción de las demandas, sería posible encontrar algún rasgo del español de América y, afortunadamente, así ocurrió. La consulta que se hizo en el *Diccionario de la lengua española* y las búsquedas de expresiones en el Corpus del español del siglo XXI (CORPES) confirmaron el hallazgo de un primer indicador para la verificación de autoría. En el segundo caso, la presencia de la hipotaxis era previsible en los tres conjuntos de textos de la muestra porque es un rasgo característico del español jurídico. No obstante, se pensó que, aun así, era posible encontrar diferentes patrones de comportamiento lingüístico que facilitaran la verificación de autoría del demandante y, en efecto, así fue. El demandante escribía menos oraciones complejas, pero más largas que las del demandado. Este patrón se repetía, además, en las demandas dubitadas. Lo mismo ocurrió con el uso más frecuente de la secuencia de palabras de 4-gramas que coincidía con las demandas dubitadas, pero no con las indubitadas del demandado.

Finalmente, el artículo ha puesto en valor una perspectiva integradora que aúna conocimientos extralingüísticos, lingüísticos, estilométricos, y estadísticos. El método empleado se puede repetir, replicar, y aporta datos objetivables acerca de la verificación de autoría. En este caso, los resultados arrojaron cinco indicadores objetivos, tres estilísticos y dos estilométricos, para la verificación de autoría del demandante. La validación objetiva es esencial porque permite al tribunal saber qué peso se debe dar a la pericial. Una perspectiva integradora allana el camino para trabajar en esa dirección.

## Bibliografía

- Alcaraz Varó, Enrique , y Hughes, Brian (2002).** *El español jurídico*, Barcelona, Ariel.
- Ainsworth, Janet (2022).** «Serving science and serving justice: ethical issues faced by forensic linguists in their role as expert witnesses», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: doing forensic linguistics*, Cham, Palgrave Macmillan, pp. 1-53, [https://doi.org/10.1007/978-3-030-84330-4\\_2](https://doi.org/10.1007/978-3-030-84330-4_2).
- Ainsworth, Janet, y Juola, Patrick (2019).** «Who wrote this?: modern forensic authorship analysis as a model for valid forensic science», *Washington University Law Review*, 96, 5, pp. 1161-1189, [https://openscholarship.wustl.edu/law\\_lawreview/vol96/iss5/10](https://openscholarship.wustl.edu/law_lawreview/vol96/iss5/10).
- Bergsma, Shane; Post, Matt, y Yarowsky, David (2012).** «Stylometric analysis of scientific articles», *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies*, pp. 327-337, <https://aclanthology.org/N12-1033.pdf>.
- Brennan, Michael; Afroz, Sadia, y Greenstadt, Rachel (2012).** «Adversarial stylometry: circumventing authorship recognition to preserve privacy and anonymity». *ACM transactions on information and system security (TISSEC)*, sección 15, 3, artículo 12, 22 páginas (en línea), <https://doi.org/10.1145/2382448.2382450>.
- Chaski, Carole (1997).** «Who wrote it? Steps toward a science of authorship identification», *National Institute of Justice Journal*, 233, 15-22, <https://www.ncjrs.gov/pdffiles/jr000233.pdf>.
- (2001). «Empirical evaluations of language-based author identification techniques», *Forensic Linguistics*, 8, 1, pp. 1-65, <https://pdfs.semanticscholar.org/2cee/ce61b8f1ae9077209fe1676e642b097bdd77.pdf>.
- (2005). «Who's at the keyboard? Authorship attribution in digital evidence investigations», *International Journal of Digital Evidence*, 4, 1, pp. 1-13, [http://www.flrchina.com/en/images/001/chaski\\_spring\\_05.pdf](http://www.flrchina.com/en/images/001/chaski_spring_05.pdf).
- (2006). «Forensic linguistics, authorship attribution, and admissibility», en Cyril H. Wecht y John T. Rago (eds.), *Forensic science and law. Investigative applications in criminal, civil, and family justice*, Boca Ratón/Londres/Nueva York, CRC Taylor y Francis, pp. 505-522.
- (2012). «Author identification in the fo-

- rensic setting», en Peter M. Tiersma y Lawrence M. Solan (eds.), *The Oxford handbook of language and law*, Oxford/Nueva York, Oxford University Press, pp. 489-504.
- (2013). «Best practices and admissibility of forensic author identification», *Journal of Law and Policy*, pp. 333-376, [http://www.ncstl.org/picture/Chaski\\_2013\\_BestPracticesAuthorship\\_BLS\\_JournalLawPolicy.pdf](http://www.ncstl.org/picture/Chaski_2013_BestPracticesAuthorship_BLS_JournalLawPolicy.pdf).
- Coulthard, Malcolm (2005)**. «Algunas aplicaciones forenses de la lingüística descriptiva», en M. Teresa Turrell (ed.), *Lingüística forense, lengua y derecho. Conceptos, métodos y aplicaciones*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra (col. Sèrie Monografies, n.º 8), pp. 249-274, <https://doi.org/10.1400/157753>.
- y **Johnson, Alison (eds.) (2010)**. *The Routledge handbook of forensic linguistics*, Londres/Nueva York, Routledge.
- Daelemans, Walter (2013)**. «Explanation in computational stylometry», *International Conference on Intelligent Text Processing and Computational Linguistics*. Berlín / Heidelberg, Springer, pp. 451-462, [https://doi.org/10.1007/978-3-642-37256-8\\_37](https://doi.org/10.1007/978-3-642-37256-8_37).
- De Jong-Lendle, Gea (2022)**. «Speaker identification», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: doing forensic linguistics*, Cham, Palgrave Macmillan, pp. 257-319, [https://doi.org/10.1007/978-3-030-84330-4\\_9](https://doi.org/10.1007/978-3-030-84330-4_9).
- Demir, Nesibe Merve (2017)**. «Better features sets for authorship attribution of short messages», *Southeast Europe Journal of Soft Computing*, 6, 1, pp. 8-12, <http://dx.doi.org/10.21533/scjournal.v6i1.128>.
- Ehrhardt, Sabine (2018)**. «Authorship attribution analysis», en Monika Rathert y Jacqueline Visconti (eds.), *Handbook of communication in the legal sphere*, Berlín/Boston, De Gruyter Mouton, pp. 169-200, <https://doi.org/10.1515/9781614514664-010>.
- Fernández-López, Mercedes (2022)**. «Expert evidence in civil law systems», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: doing forensic linguistics*, Cham, Palgrave Macmillan, pp. 85-104, <https://doi.org/10.1007/978-3-030-84330-4>.
- Fobbe, Eilika (2022)**. «Authorship identification», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: doing forensic linguistics*, Cham, Palgrave Macmillan, pp. 185-217, [https://doi.org/10.1007/978-3-030-84330-4\\_7](https://doi.org/10.1007/978-3-030-84330-4_7).
- Gil-Salom, Luz; Soler-Monreal, Carmen; Stuart, Keith; y Candela-Coderch, Jorge (2004)**. *TextWorks*, Valencia, Universidad Politécnica de Valencia (Departamento de Idiomas).
- Grant, Tim (2007)**. «Quantifying evidence in forensic authorship analysis». *International Journal of Speech and the Law*, 14, 1, pp. 1-25, <https://doi.org/10.1558/ijssl.v14i1.1>.
- (2008). «Approaching questions in forensic authorship analysis», en John Gibbons y M. Teresa Turrell (eds.), *Dimensions of forensic linguistics*, Ámsterdam/Filadelfia, John Benjamins, pp. 215-229.
- Guillén-Nieto, Victoria (2020)**. «The relevance of context in plagiarism detection: The case of a professional legal genre», *Ibérica*, 40, pp. 101-122.
- y **Stein, Dieter (eds.) (2022)**. *Language as evidence: doing forensic linguistics*, Cham, Palgrave Macmillan, <https://doi.org/10.1007/978-3-030-84330-4>.

- org/10.1007/978-3-030-84330-4.
- (2022). «Introduction: theory and practice in forensic linguistics», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as Evidence. Doing Forensic Linguistics*, Cham, Palgrave Macmillan, pp. 1-33, [https://doi.org/10.1007/978-3-030-84330-4\\_1](https://doi.org/10.1007/978-3-030-84330-4_1).
- Halvani, Oren; Winter, Christian, y Pflug, Anika (2016)**. «Authorship verification for different languages, genres and topics», *Digital Investigation*, 16, pp. 33-43, <https://doi.org/10.1016/j.diin.2016.01.006>.
- Hammel, Andrew (2022)**. «Linguistic expert evidence in the common law», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: Doing forensic linguistics*, Cham, Palgrave Macmillan, pp. 55-84, [https://doi.org/10.1007/978-3-030-84330-4\\_3](https://doi.org/10.1007/978-3-030-84330-4_3).
- Holmes, David I. (1998)**. «The evolution of stylometry in humanities scholarship», *Literary and Linguistic Computing*, 13, pp. 111-117, <https://doi.org/10.1093/llc/13.3.111>.
- Houvardas John, y Stamatos, Efstathios (2006)**. «N-gram feature selection for authorship identification», en Jérôme Euzenat y John Domingue (eds.), *Artificial intelligence: Methodology, systems, and applications*. AIMS. *Lecture notes in computer science*, Berlín/Heidelberg, Springer, 4183, pp. 77-86, [https://doi.org/10.1007/11861461\\_10](https://doi.org/10.1007/11861461_10).
- Kaczynski, Theodore (1995)**. *The UNABOMBER Manifesto*, <https://archive.nytimes.com/www.nytimes.com/library/national/unabom-manifesto-1.html>.
- Kniffka, Hannes (2000)**. «Anonymous authorship analysis without comparison data? A case study with methodological implications», *Linguistische Berichte*, 182, pp. 179-198.
- Kredens, Krzysztof, y Coulthard, Malcolm (2012)**. «Corpus linguistics and authorship identification», en Peter Tiersma y Lawrence M. Solan (eds.), *The Oxford Handbook of Language and Law*, Oxford/Nueva York, Oxford University Press, pp. 504-516.
- Ley de Propiedad Intelectual 1/1996 de 12 de abril de 1996**. <https://www.boe.es/eli/es/rdlg/1996/04/12/1>.
- López-Escobedo, Fernanda; Sierra, Gerardo, y Solórzano, Julián (2019)**. «SAUTEE: un recurso en línea para análisis estilométricos». *LinguaMática*, 11, 1, pp. 69-81, <https://doi.org/10.21814/lm.11.1.270>.
- Real Academia Española (2014)**. *Diccionario de la lengua española* [en línea], 23.ª edición, versión 23.6. <https://dle.rae.es/>.
- (2021). *Banco de datos (CORPES XXI) v. beta 0.94* [en línea]. *Corpus del Español del Siglo XXI (CORPES)*, <http://www.rae.es>.
- Love, Harold (2002)**. *Attributing Authorship: An Introduction*, Cambridge, Cambridge University Press.
- Stamatos, Efstathios; Daelemans, Walter; Verhoeven, Ben, y Potthast, Martin (2014)**. Overview of the author identification task at PAN 2014, *Working Notes for CLEF 2014 Conference*, Sheffield, UK, septiembre 15-18, pp. 877-897, <http://ceur-ws.org/Vol-1391/inv-pap3-CR.pdf>.
- Stamatos, Efstathios (2009)**. «A survey of modern authorship attribution methods», *Journal of the American Society for Information Science and Technology*, 6, 3, pp. 538-556, <https://doi.org/10.1002/asi.21001>.
- Svartvik, Jan (1968)**. «The Evans statements. A case for forensic linguistics», en Alvar Ellegard (ed.), *Acta Universitatis Gothoburgensis. Gothengurg Studies in English*, 20, 2 vols., Gotemburgo,

- Elanders Boktryckeri Aktiebolag, <https://www.thetext.co.uk/Evans%20Statements%20Part%201.pdf>.
- Tiersma, Peter M., y Solan, Lawrence M. (eds.) (2012).** *The Oxford handbook of language and law*, Oxford/Nueva York, Oxford University Press.
- Turchie, Terry D. (1996).** *Affidavit of Assistant Special Agent in Charge*. U. S. District of Montana, <http://www.courtvt.com/archive/casefiles/unabomber/documents/affidavit.html>.
- van Dijk, Teun A. (2008).** «Context theory and the foundation of pragmatics», *Studies in Pragmatics*, 10, pp. 1-13.
- van Halteren, Hans (2022).** «Automatic authorship investigation», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: doing Forensic Linguistics*, 219-255, Cham, Palgrave Macmillan, pp. 219-255, [https://doi.org/10.1007/978-3-030-84330-4\\_8](https://doi.org/10.1007/978-3-030-84330-4_8).
- Zaško-Zielińska, Monika (2022).** «The linguistic analysis of suicide notes», en Victoria Guillén-Nieto y Dieter Stein (eds.), *Language as evidence: doing forensic linguistics*, Cham, Palgrave Macmillan, pp. 373-417, [https://doi.org/10.1007/978-3-030-84330-4\\_11](https://doi.org/10.1007/978-3-030-84330-4_11).