

# La imitación como contraargumento en peritajes de atribución de autoría: estudio de un caso

Forgery as counterargument in authorship attribution reports: A case study

Sheila Queralt

Universitat de Barcelona, Laboratorio sq-Lingüistas Forenses

Roser Giménez García

Universitat de Barcelona, Laboratorio sq-Lingüistas Forenses

## Resumen

Este artículo tiene como objetivo probar el supuesto de que el estilo de un autor en textos escritos no solo es único, sino también irrepetible. El marco teórico apoya este supuesto e indica que se requeriría un gran dominio lingüístico y la capacidad de sostener un alto grado de atención consciente, los cuales son muy difíciles de alcanzar. El estudio se basa en el material de un caso real de chantaje a un funcionario público, en el que participó una de las autoras, así como un corpus de textos producidos con el fin de imitar el estilo del funcionario. Se empleó una combinación de análisis cualitativo para la identificación

de estructuras marcadas y prominentes con un análisis cuantitativo de estadística descriptiva. Se observó que los imitadores únicamente reproducen los rasgos más visibles y no son capaces de identificar e incorporar todos los atributos de la escritura del autor original en sus textos. Además, en el caso de detectar una marca particular, no la reproducen con la misma frecuencia. Los resultados permiten refutar la hipótesis sobre la posibilidad de la imitación del estilo de un autor de forma suficientemente meticulosa y consistente para evitar ser detectado por un lingüista forense.

Palabras clave: lingüística forense; evidencia lingüística; peritaje lingüístico; suplantación; estilo de un escritor

## Abstract

The aim of this paper is to scientifically test the assumption that the writing style of an author is not only unique but unrepeatable. The literature supports this assumption and suggests that successful forgery would require both the mastery of language and a sustained level of conscious attention, which are extremely difficult to attain. Our study used a corpus of texts by and material from a real blackmail case in which one of the authors worked. The study combines a qualitative analysis for the identification of marked and prominent structures combined with quantitative descriptive statis-

tics. The results lead us to reject the hypothesis that the writing style of an author can be successfully forged. Forgers can replicate the most visually prominent traits, but they are unable to identify and incorporate all the traits of the original author in their productions. Moreover, when they do identify a trait, forgers are unable to replicate it with the same frequency as the original author. We therefore conclude that the writing style of an author cannot be forged as thoroughly and consistently as needed be for the forger to remain unnoticed by forensic linguistic analysis.

**Keywords:** forensic linguistics; linguistic evidence; linguistic expert report; authorship disguise; author's style

Fecha de recepción del artículo: 24 de octubre de 2017

Fecha de recepción de la versión revisada: 14 de junio de 2018

Fecha de aceptación: 19 de junio de 2018

La correspondencia relacionada con este artículo debe dirigirse a:

Sheila Queralt

sheila.queralt@clicenciats.cat

## 1. Introducción

El auge en el uso de plataformas electrónicas para comunicarse, como Twitter, WhatsApp, Facebook, Instagram, y el incremento de *softwares* para anonimizar y dificultar el rastreo de identidades por internet se están reflejando en la tipología de periciales presentadas en los tribunales. Por un lado, está disminuyendo el número de periciales caligráficas e informáticas en los juzgados y, por otro, está aumentando el número de periciales lingüísticas, ya que estas últimas no analizan la forma manuscrita del mensaje ni tampoco rastrean su origen, sino que se basan en el análisis de la estructura del lenguaje usada por el autor de un texto anónimo.

De igual modo, cada vez es más común la intervención de un perito lingüista forense en casos judiciales en España y el extranjero para asesorar al juez sobre la autoría de varios escritos. En general, el análisis de autoría se ha definido como aquel proceso mediante el cual se examinan las características lingüísticas de un texto para extraer conclusiones sobre su autoría (Zheng, Qin, Huang & Chen, 2003). El lingüista debe determinar la probabilidad de que un autor haya escrito un texto o, todo lo contrario, la probabilidad de que no sea el autor del mismo. Para ello, puede servirse de métodos cualitativos y/o cuantitativos, dependiendo de la extensión y calidad de los textos comparados. En muchas de las ocasiones en que el perito determina la probabilidad de que el autor sugerido haya producido también los textos de autoría desconocida, la parte contraria utiliza como argumento principal la hipótesis de un segundo autor capaz de reproducir el estilo del acusado. Por ejemplo, en el caso de una empresa, otros trabajadores de la misma sección o, en el caso de un matrimonio, el otro miembro de la pareja.

Hasta el momento, los lingüistas forenses defienden las conclusiones sobre la autoría de un escrito, a las que llegan mediante métodos cualitativos o cuantitativos, de análisis manual o computacional, con base en los principios teóricos utilizados en lingüística forense. El objetivo de este estudio es servir como primer paso en el análisis de la autoría para poder refutar de forma empírica el argumento recurrente de la parte contraria: la hipótesis a favor de la posibilidad de la imitación del estilo de un autor y, por tanto, la exculpación del acusado.

## 2. Presentación del caso real

Con el fin de poder argumentar la dificultad de reproducir el estilo de otro autor sin ser detectado, se ha llevado a cabo una investigación a partir de un caso real de chantaje a un funcionario en el que se formuló esta hipótesis, en la cual se basó gran parte de la defensa.<sup>1</sup>

El caso sucedió en el año 2012 en España entre un político y un alto directivo de una entidad bancaria. En la Figura 1 se presenta la sucesión de los hechos.

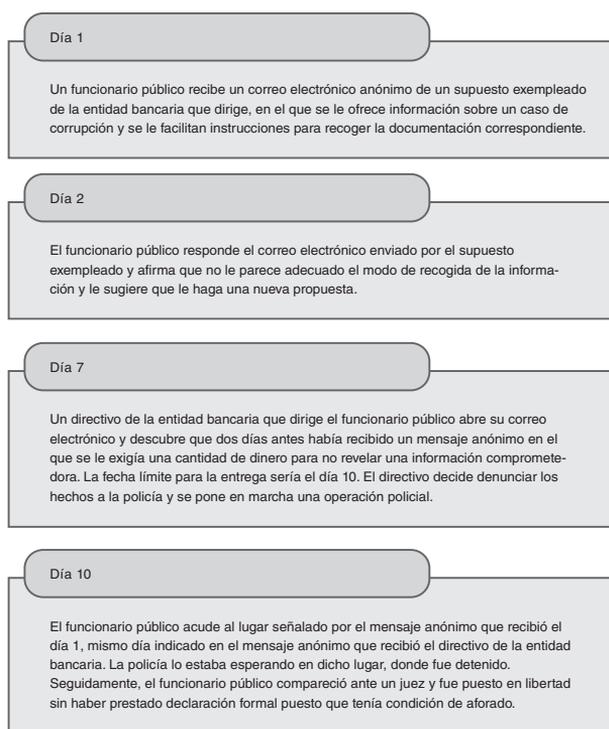


FIGURA 1. Sucesión de los hechos del caso real

<sup>1</sup> Las referencias personales o contextuales han sido anonimizadas. La información sobre el caso, así como los anónimos recibidos, están disponibles al público. (Revelación: Una de las autoras del trabajo participó en este caso como perito lingüista forense de parte.)

Tanto el magistrado-juez instructor como el demandante, el directivo de una institución bancaria, solicitan un peritaje lingüístico de atribución de autoría en el que se debía comparar el texto anónimo recibido por el directivo con un corpus de textos del principal sospechoso, el funcionario público, y ahora imputado. Además, durante el proceso de instrucción, el magistrado-juez instructor recabó indicios que sugerían que la persona que redactó y envió dicho mensaje anónimo pudiera ser el imputado del caso, es decir, el funcionario público. Así pues, uno de los aspectos a considerar en el juicio era si el texto en cuestión podría atribuirse o no al imputado.

El imputado sugirió en diversas comunicaciones públicas y privadas una versión distinta de los hechos, según la cual él sería víctima de una conspiración que trataría de involucrarlo en el asunto para desprestigiarlo políticamente.

### 3. Fundamentos teóricos de la atribución de autoría

Este trabajo se apoya en el principio de *individualidad lingüística* (Johnstone, 2000; Chaski, 2001; McMenamin, 2001: 93–97; Coulthard, 2004; Coupland, 2007; Howald, 2009; Grant, 2010), el cual postula que cada individuo se expresa mediante un *estilo idiolectal* (Turell, 2010) característico formado por la conjunción de sus cualidades sociolingüísticas. La asunción de la hipótesis del estilo es fundamental para poder establecer una medida de similitud idiolectal que pueda determinar la probabilidad de que dos muestras lingüísticas hayan sido producidas o no por el mismo autor (Queralt, 2014). Esta premisa teórica es ampliamente aceptada por la comunidad de lingüistas forenses de todo el mundo como el principio básico para lidiar con el problema de atribución de autoría, a pesar de que, desde un punto de vista teórico, no pueda considerarse todavía completamente consolidada en todos sus términos (Grant, 2010). Según esta premisa, un autor reproduce de forma sistemática en sus escritos ciertos rasgos idiosincrásicos, característicos y distintivos. La defensa de la parte contraria argumenta que dichos rasgos podrían ser susceptibles de ser identificados de forma consciente por el imitador y que, por lo tanto, podría reproducirlos de manera intencional con mayor o menor grado de similitud.

Las características específicas y únicas de cada autor, que conforman su estilo idiolectal, se definen mediante los conceptos complementarios de *marcadez* y

*prominencia*. La noción de marcadez, en el sentido utilizado en lingüística forense, fue propuesta por Jakobson (1971 [1956]). Las formas marcadas en el discurso de una persona aportan información adicional y específica más precisa que las formas no marcadas, ya que conforman elecciones poco comunes en su producción. Una forma marcada será por tanto considerada una característica idiolectal, por tratarse de un elemento lingüístico no esperado o poco frecuente que se puede dar, por ejemplo, por contacto de lenguas. Un ejemplo sería el uso de un elemento léxico concreto en una posición en la que se esperaría uno distinto, según el comportamiento y/o características de su autor.

Por otro lado, la noción de prominencia descrita por Butler (1998) y Abecassis (2002) acerca el análisis del discurso a la lingüística de corpus, ya que se construye con base en todas las unidades lingüísticas que destacan estadísticamente en un corpus. Las características prominentes de un estilo idiolectal son aquellas que coinciden con más frecuencia en las producciones de un autor que en las de otros. Por lo tanto, al contrario que la marcadez, la prominencia se observa en el uso reiterado y frecuente de un conjunto único de características específicas que definen un estilo idiolectal.

Recientemente, una corriente de expertos en lingüística computacional y ciencias de la computación han desarrollado y defendido estudios en que se utilizan medidas estilométricas producidas por *software* especializado en la comparación y medición de textos (véase por ejemplo El Bouanani & Kassou, 2014, o Stamatatos, 2009). No obstante, desde un punto de vista lingüístico, muchos de estos estudios presentan problemas metodológicos, ya que es difícil proporcionar una explicación acerca de la capacidad distintiva de las medidas empleadas para clasificar textos según las teorías lingüísticas actuales del idiolecto y de la variación lingüística (Wright, 2014: 21). Un ejemplo muy conocido de un método completamente computacional e inexplicable desde la lingüística, cuya eficacia fue puesta a prueba y rechazada por Canter y Chester (1997), es el método Cusum (Coulthard & Johnson, 2007: 167–169). Este método, que fue desarrollado por Morton (1991), utiliza la oración como unidad básica de medida y calcula la frecuencia de aparición de variables como el número de sustantivos o de palabras de dos y tres letras en cada oración. Los resultados parten de la suma acumulativa de la desviación respecto de la media para todo el texto del total de instancias de la variable en cuestión y del total de palabras por oración.

Antes de que varios estudios (Sandford, Aked, Moxey & Mullin, 1994, o Canter & Chester, 1997) demostrasen que este método no era capaz, como se había defendido, de detectar inserciones de otros textos en los textos analizados, se utilizó al menos en cuatro ocasiones en juzgados del Reino Unido. Este recurso ejemplariza el riesgo de delegar completamente en ecuaciones o *softwares* la resolución de pruebas de autoría por suponer que su automaticidad y capacidad de ser generalizables garantiza su eficacia y fiabilidad. Además, un inconveniente que presenta el uso de estos métodos es que la lectura de los resultados y gráficos proporcionados por *softwares* diseñados para medir la similitud entre distintos textos no siempre es indiscutida, incluso entre lingüistas computacionales (Somers, s. f.). Otra limitación de los métodos computacionales está relacionada con el hecho de que, generalmente, es difícil aplicar medidas con un nivel de acierto satisfactorio a través de varios géneros textuales, ya que se diseñan y prueban específicamente para ciertos géneros (Stamatatos, 2009). Sin embargo, dichas características de diseño son opuestas a la realidad de la lingüística forense, en que es habitual contar con muestras lingüísticas pertenecientes a géneros diferentes (correos electrónicos en el trabajo, cartas de extorsión, publicaciones familiares en redes sociales, etc.). Además, los corpus comparados en esta disciplina son a menudo de extensiones reducidas. Por el contrario, un requerimiento generalizado en estudios computacionales de atribución de autoría es el uso de corpus muy extensos (Stamatatos, 2009). Por último, cabe destacar que el número de métodos y medidas computacionales con unos niveles de acierto y fiabilidad suficientes como para ser utilizados en casos judiciales es todavía escaso. Así pues, teniendo en cuenta los inconvenientes de los métodos automáticos reseñados, en el presente estudio se opta por métodos de análisis cualitativos y de estadística descriptiva para identificar y evaluar las características lingüísticas imitadas por los participantes.

Según Bernárdez el texto contiene “una serie de rasgos, de peculiaridades lingüísticas en todos los niveles, desde el más superficial al más profundo, desde el más estricto (fonología y morfología) al más libre (organización textual, relaciones semántico-conceptuales) pasando por los intermedios (sintaxis)” (2003: 16). De este modo nuestro estudio pretende analizar características de estos niveles para observar su reproducibilidad siguiendo también la perspectiva de la teoría de la relevancia de Wilson y Sperber (1993), utilizada entre otros por Regina Blass

(1990), quien propone concebir el texto como un conjunto de instrucciones para que el receptor pueda (re)construir el texto. En nuestro estudio, la reconstrucción del texto tiene como objetivo que el receptor adapte las instrucciones del emisor y reconstruya un texto a su semejanza. La imitación aparentemente conlleva operaciones cognitivas complejas e insuficientemente comprendidas (Heyes, 1993) que involucran la visión, la percepción, la representación, la memoria y el control motor (Borenstein & Ruppín, 2005).

La imitación, sus consecuencias para el aprendizaje de comportamientos sociales y los procesos cognitivos que la posibilitan conforman los objetivos de numerosos estudios empíricos (*e. g.*, Iacoboni, 2009). Aun así, las operaciones cognitivas ligadas específicamente a la reproducción e imitación de estilos textuales siguen siendo un área por lo general desconocida y que requiere investigación conjunta por parte, entre otros, de psicólogos, neurocientíficos y lingüistas. Sin embargo, en la intersección entre estudios sobre la cognición y el aprendizaje de idiomas sí encontramos investigaciones cuyos descubrimientos pueden contribuir a nuestros objetivos. De especial relevancia para entender cómo un autor puede imitar el estilo idiolectal de otro con éxito son los artículos científicos que buscan respaldar propuestas pedagógicas de la escritura de ensayos mediante el análisis e imitación de modelos textuales considerados “correctos” (*e. g.*, Geist, 2005), así como aquellos que se proponen explicar la adquisición lingüística de niños y estudiantes de lenguas extranjeras (*e. g.*, Adank, Hagoort & Bekkering, 2010).

En primer lugar se ha postulado que, para poder imitar, el autor tendría que realizar una lectura consciente y observar de forma precisa el comportamiento del otro (Arriaga, Ortega, Meza, Huichán, Juárez, Rodríguez & Cruz, 2006: 88). Asimismo, tendría que seguir las instrucciones de la teoría de la relevancia y analizar las distintas estructuras del emisor. Los análisis de las estructuras van a permitir que los lectores/autores expandan “their understanding of how an author in a particular context will use language to appeal to a particular audience”<sup>2</sup> (Foster, 1989: 18).

<sup>2</sup> “[S]u comprensión de cómo un autor usará la lengua en un contexto concreto para atraer a un público determinado”. Las traducciones son nuestras.

De este modo, se puede determinar que el proceso de imitación es una actividad consciente, es decir, “imitation is rooted in the speaker’s perception of another speaker or variety (or folk artifact of a variety) and reflects their awareness of it”<sup>3</sup> (Brunner, 2011: 24) y que, por lo tanto, para ser capaz de imitar “one must be aware of what is to be imitated and how it can be imitated. Thus, imitation is closely connected with observation and analysis (‘exploration’ as Piaget calls it)”<sup>4</sup> (Geist, 2005: 172).

En este estudio seguimos los supuestos revisados en el marco teórico que se utilizan generalmente para analizar la imitación consciente, que implican verificar la hipótesis de quién ha escrito un texto y qué elementos pueden representar el estilo único del autor, tal y como indica Oakman (1984: 140). Por lo tanto, como el autor ha querido de manera deliberada imitar el estilo del otro autor, es esperable que, si el ejercicio ha sido exitoso, el texto resultante comparta características identificables y marcadas del autor imitado (Irizarry, 1989: 228).

Como se desprende de la base teórica descrita en este apartado, cabe suponer, y de hecho se pretende demostrar que, aunque el imitador intente identificar los rasgos y reproducirlos, difícilmente será capaz de identificar todos los rasgos lingüísticos analizados en todos los niveles ni podrá evitar incluir en el texto suplantador rasgos de su propia individualidad lingüística.

#### 4. Objetivos e hipótesis del estudio

Durante las vistas orales<sup>5</sup> se plantearon distintas hipótesis acerca de la autoría del texto anónimo, en lo que concierne a la plausibilidad de la imitación, de estilos compartidos e, incluso, del contagio de estilos. En concreto, el letrado de la defensa y el magistrado-juez instructor plantearon una y dos cuestiones, respectivamente, que se reproducen a continuación de forma literal.

<sup>3</sup> “[L]a imitación está enraizada en la percepción de un hablante de otro hablante o variedad (o artefacto cultural de una variedad) y refleja la conciencia de ese hablante o variedad”.

<sup>4</sup> “[S]e debe ser consciente de lo que se va a imitar y de cómo se puede imitar. Así pues, la imitación está estrechamente relacionada con la observación y el análisis (lo que Piaget denomina ‘exploración’).”.

<sup>5</sup> Vista oral: fase en la que se practican las pruebas y efectúan las alegaciones en presencia del juez para dictar sentencia.

a) Pregunta del letrado de la defensa:

A la hora de escribir, cuando leemos un texto varias veces e intentamos explicar ese texto, ¿podemos copiar parte de este texto? ¿podemos coger e imbuirnos de este texto?

b) Preguntas del magistrado-juez instructor:

¿Es posible que una tercera persona que conociera el estilo del imputado imitara ese estilo para hacer creer que era él el autor de ese texto?

¿Puede ocurrir que, en función de la educación, de todas las circunstancias que conforman la experiencia vital de una persona (cultura, estudios, relaciones personales) haya dos personas que escriban de la misma manera o de forma tan parecida que se pueda confundir las escrituras de ambos?

#### 4.1. *Objetivo*

Este estudio conforma una primera aproximación para refutar de forma empírica el argumento recurrente de la parte contraria de que se pudiera reproducir el estilo de un autor en un mensaje similar al que motiva el procedimiento, con un grado de similitud tal que la suplantación no pudiera ser detectada mediante las técnicas habitualmente aplicadas por la lingüística forense. Con este fin, se diseñó una prueba de imitación en la cual se pidió a treinta personas con perfiles de edad y grupo social parecidos que redactaran una carta de amenazas intentando reproducir “el estilo” del imputado, del que se proporcionaban cuatro textos extraídos de su blog (de ahora en adelante, corpus indubitado), con un total de 1836 palabras. Estos textos y los que debían producir los participantes, pues, pertenecen a géneros diferentes. Sin embargo, como suele suceder en casos de atribución de autoría, no fue posible obtener producciones del imputado más fácilmente comparables con las cartas de las amenazas. Del mismo modo, la extensión del corpus indubitado es limitada, aunque, como recordaba Coulthard (1994: 28), siempre habrá discusiones acerca del tamaño idóneo de una muestra que se pueda considerar verdaderamente representativa. Aun así,

de forma orientativa, se puede observar que el corpus indubitado supera el mínimo recuento de palabras aconsejable en caso de comparación forense de textos escritos para determinar si existe o no caso lingüístico (Queralt, 2014: 37).

En concreto, se pretende analizar:

- a) cuáles son las variables lingüísticas del corpus proporcionado a los participantes que les son más fáciles de identificar e imitar, tratándose de hablantes de la misma lengua y
- b) cuáles son las variables lingüísticas del corpus que les han resultado más difíciles de imitar.

#### 4.2. *Hipótesis*

Las hipótesis de trabajo son las siguientes:

- 1) El proceso de imitación es una actividad cognitiva consciente e intencional que requiere un buen dominio lingüístico, de una buena capacidad para la lectura consciente y para la atención a la forma, y de una cierta capacidad para el control metacognitivo de la propia escritura.
- 2) La forma de expresión de cada persona es resultante de la combinación de múltiples características, de modo que, para determinar los rasgos idiosincrásicos del estilo de un autor, habrá que considerar un amplio espectro de niveles y variables de la lengua.
- 3) Es probable que, en función de la intuición y conocimientos del imitador, se capten y trasladen únicamente determinados conjuntos de rasgos, y que estos queden impregnados inevitablemente de su propio estilo personal.

#### 4.3. *Corpus indubitado*

Los textos indubitados (TI) son cuatro textos extraídos del blog del imputado escritos en cuatro fechas diferentes del año 2012. Los cuatro textos pertenecen al género de artículos de opinión y se detallan en el Cuadro 1.

CUADRO 1. Corpus indubitado

TÍTULO ARTÍCULO	NÚMERO DE PALABRAS
I1	389
I2	341
I3	821
I4	285

Se trata de un blog en el que, presumiblemente, el imputado ha escrito a título personal, y constituye la única referencia disponible en internet de textos de su autoría. Por lo tanto, cabe pensar que, siendo esta la única fuente accesible de manera directa, una persona que pretendiera simular el estilo de su autor acudiría muy probablemente a ella como punto de referencia.

#### 4.4. *Corpus de simulación*

En este estudio usamos el criterio simplificado de  $n > 30$  como justificación para el supuesto del tamaño de la muestra requerido, ya que no es posible identificar una extensión de muestra mínima específica que sea suficiente para todos los casos (Triola, 2004: 319) y, en particular, disponemos de una distribución de la población normal. Por lo tanto, el corpus de simulación está compuesto por 30 cartas de extorsión simuladas por varones españoles adultos con una media de edad de 43.6 ( $DE = 6.13$ ) con estudios universitarios en los que se tratan aspectos lingüísticos (filología, traducción, periodismo y derecho). Podría ser que la formación en lingüística de los participantes repercutiera positivamente en la identificación de variables imitables del corpus indubitado, aunque no disponemos todavía de estudios comparativos que puedan apoyar esta suposición.

### 5. Metodología

En este estudio se ha llevado a cabo un análisis cualitativo y un análisis cuantitativo. El análisis cualitativo ha consistido en un examen lingüístico minucioso sobre los textos indubitados para detectar los rasgos marcados y prominentes que puedan revelar posibles variables idiosincrásicas del autor. Seguidamente, se ha realizado una comparación de esos rasgos con el corpus de textos de simu-

lación. La comparación ha sido de forma manual, teniendo en cuenta las características lingüísticas y textuales que se asemejan o difieren. Además, los textos han sido procesados de manera automática para detectar el vocabulario y las estructuras compartidas de manera literal con el *software* de detección de plagio, CopyCatchGold de CFL Software, creado por David Woolls.

Con el fin de valorar hasta qué punto los participantes son capaces de imitar el estilo del autor del texto indubitado, se ha realizado un análisis cuantitativo mediante estadística descriptiva. Para ciertas variables (en concreto, el uso de léxico especializado propio de la medicina y el uso reiterado de frases hechas y de adjetivos antepuestos), no ha sido posible comparar las frecuencias observadas en los corpus estudiados con ningún corpus de referencia o estudio que pueda aportar datos sobre cuál sería su frecuencia esperable.

Actualmente, no disponemos de estudios que puedan responder a cuestiones cómo: “¿cuántas realizaciones de la variable *adjetivo antepuesto* se observan de media en las producciones escritas en español?” o “¿con qué frecuencia encontramos léxico especializado de medicina en textos sobre política?” Así pues, no ha sido posible obtener ninguna medida cuantitativa que pueda orientarnos sobre su grado de marcadez y prominencia, más allá de los conocimientos lingüísticos de las autoras como expertas y hablantes de la lengua. Las dificultades de obtener este tipo de información se deben, entre otras, a la naturaleza de las variables (no tratamos de calcular la marcadez de cierta variable, sino la marcadez de una frecuencia concreta de la variable en un género textual específico). Además, el hecho de que haya pocos estudios sobre el género textual analizado, las cartas de extorsión, así como la falta de disponibilidad de un corpus de referencia específicamente diseñado para el estudio de producciones en contextos forenses, que recoja un número suficiente de textos, contribuye a la imposibilidad de calcular la referencia de distribución poblacional (en inglés, *base rate knowledge*) (Grant, 2007) de este tipo de variables, ya que el uso de corpus de referencia de la lengua generales —como el Corpus de Referencia del Español Actual (CREA) o el Corpus del Español del Siglo XXI (CORPES)— no proporciona los datos requeridos para calcular la distribución poblacional o frecuencia esperable de dichas variables.

Las variables seleccionadas para este estudio y, por tanto, idiosincrásicas del corpus indubitado, han sido las que destacaban por su prominencia: detecta-

das de forma cualitativa teniendo en cuenta frecuencias altas de variables no esperables como, por ejemplo, el uso de metáforas sobre el campo de la medicina en un texto sobre política; o su marcadez: es decir, variables que aparecían en el corpus indubitado en una variante marcada como podrían ser posiciones antepuestas o pospuestas de una variable o la preferencia de una variante de una variable polimórfica.

Una vez seleccionadas las variables más idiosincrásicas se ha procedido a su análisis en el corpus de simulación para determinar si el imitador era capaz de reproducirla con éxito en su escrito. En primer lugar, el autor debía ser capaz de detectar la realización idiosincrásica de cada variable analizada presente en el corpus indubitado y, en segundo lugar, de imitarla. Se consideraba que el autor reproducía con éxito esa variable si la introducía en su escrito y, en caso de que la variable presentara variantes, si reproducía la variante más utilizada por el autor original.

Para obtener el porcentaje de éxito en la imitación de cada variable, se ha calculado cuántos autores del corpus de simulación eran capaces de reproducir de forma exitosa la variable. Cuantos más autores son capaces de imitar la variable, más alto es el porcentaje de éxito en la imitación y, por tanto, se ha podido determinar qué variables lingüísticas del corpus indubitado son más fáciles y más difíciles de ser imitadas por hablantes de la misma lengua.

## 6. Variables

El conjunto de variables que se analiza en una pericial de comparación forense de textos escritos no se limita exclusivamente al listado de variables estudiadas en este artículo, sino que, por lo contrario, existen muchas otras en función del género discursivo, el número de muestras o la casuística del caso. Sin embargo, para fines de este estudio se acotaron las variables y se analizaron las que más se suelen utilizar, independientemente del contexto de estudio (véase Turell, 2011; Queralt, 2014; Hernández, 2016).

De este modo, tal y como se detalla en el Cuadro 2, en este estudio se analizan cuatro módulos lingüísticos: léxico-semántico, morfosintáctico, pragmático-discursivo y complejidad. En primer lugar, se encuentra el módulo léxico-semántico, en el que se analiza el uso de léxico especializado e infrecuente,

de anglicismos, de frases metafóricas, así como frases hechas y expresiones. A nivel morfosintáctico, se analiza el orden de palabras, el uso de tiempos verbales y los pronombres relativos. En el grupo de variables pragmático-discursivas, se analizan los marcadores discursivos, la expresión del énfasis y el formato de incisos. Por último, como variables de complejidad, se analiza el número de palabras por frase.

CUADRO 2. Variables idiosincrásicas del corpus indubitado analizadas

MÓDULO LINGÜÍSTICO	VARIABLES IDIOSINCRÁSICAS DEL CORPUS INDUBITADO
Léxico-semántico	Léxico especializado
	Palabras poco frecuentes
	Anglicismos
	Frases hechas
	Metáforas
	Expresiones (estructuras propias del autor)
Morfosintáctico	Orden de palabras: adjetivo antepuesto
	Perífrasis de obligación
	Imperfecto de subjuntivo
	Futuro
	Pronombres relativos
Pragmático-discursivo	Marcadores discursivos en posición inicial
	Expresión del énfasis
	Formato de incisos
Complejidad	Media de palabras por frase

## 7. Análisis comparativo de las variables presentes en el texto indubitado y en los textos imitativos

### 7.1. *Léxico-semántico*

El autor de los textos indubitados realiza un uso prominente de cuatro variables léxico-semánticas: uso de vocabulario y metáforas del campo de la medicina,

uso de frases hechas y uso de anglicismos. También destaca por el empleo de palabras poco frecuentes en español y, por tanto, marcadas.

### 7.1.1. Léxico especializado

El autor del conjunto de los textos indubitados recurre a léxico propio y metáforas relacionadas con la medicina para explicar cuestiones sociopolíticas y también del campo léxico de la política de finanzas. Esta variable prominente ha sido detectada por la mayoría de los imitadores y se podría explicar, presumiblemente, por la formación universitaria en medicina del autor. En algunos casos, el imitador ha reproducido las mismas metáforas y, en otros, se ha plasmado el mismo vocabulario, pero modificando las expresiones. Se muestran algunos ejemplos en el Cuadro 3.

CUADRO 3. Ejemplos de léxico y expresiones de campos especializados

	CORPUS INDUBITADO	CORPUS SIMULACIÓN
Léxico especializado del campo de la medicina	Bisturí	Me produce el mismo sentimiento que el de un médico que es consciente de que será peor el remedio que la enfermedad —que en este caso la enfermedad parece haber calado hasta lo más hondo de la sociedad— [1]
	Ha salido pus, el empiema estaba ahí aunque algunos lo quisieran negar. Toca cirugía y no analgesia.	No dispondré de calmantes para su dolor. [1]
	Un sistema, el financiero, aquejado de la gangrena de la corrupción.	Se producirá una hemorragia en el sistema financiero. [2]
	En contra de sus pulsiones.	De nuevo se producirá una hemorragia en el sistema [2]
	Los médicos llamamos “shock hipervolémico” a la situación en la que el enfermo puede morir como consecuencia de la pérdida de flujo sanguíneo, por hemorragia.	la gangrena no se cura con analgésicos los médicos sabemos que la gangrena no se cura con analgésicos sino con una operación quirúrgica [4]
	No se puede poner al paciente a hacer ejercicio si no tiene sangre en las arterias.	hay que aislar y extirpar la gangrena [4]
	Se haya decidido a sajar con mayor profundidad del	sustenta i apoya la gangrena [sic] [4]
	parecía que nuestro país recuperaba las constantes vitales [5]	

CUADRO 3. Ejemplos de léxico y expresiones de campos especializados (*continuación*)

	CORPUS INDUBITADO	CORPUS SIMULACIÓN
Léxico especializado del campo de la medicina		En tus manos está pues que la reunión de ayer no tenga que generar unos efectos quirúrgicos irreversibles para tus intereses [8]  provocar una amnesia permanente [8]  vacunarse frente a posibles represalias [9]
Léxico especializado del campo de la política de finanzas	sistema financiero disponibilidad de crédito inflábamos la mayor burbuja inmobiliaria reformas financieras presupuesto público ajustes en su dispendio pago de intereses políticas de estímulo económico créditos suficientes para operar actividad bancaria	sistema financiero óptimo [6] saqueo de la lógica financiera [6] subvenciones públicas [6] inversiones [17] estimular la economía [17] inflaron los honorarios [17] institución financiera [17] gestiones ilícitas [11] facturación de Caja Navarra [11] malversación de fondos [19] expolio de la entidad [19] actividad ilícita dentro de la entidad [19]

### 7.1.2. Palabras poco frecuentes

El autor de los textos indubitados utiliza palabras *marcadas*, es decir, palabras con una baja frecuencia respecto a sus sinónimos según datos de corpus. Además, el autor alterna el uso de términos cultos como *verbigracia* y términos propios de registros más coloquiales como *esquilmar* o *arramplar*.

La frecuencia de aparición de estos términos se ha procesado mediante la lingüística de corpus analizando los usados por el autor de los textos indubitados y algunos sinónimos en el CORPES XXI, realizado por la Asociación de Academias de la Lengua Española (Asale) y que recoge aproximadamente 200 millones de formas. Aun así, es importante destacar una vez más que dicho corpus de referencia solo puede proporcionar datos orientativos, ya que, a pesar de su extensión, los textos que recoge son de índole muy diversa y diferente a los corpus estudiados aquí.

Como se observa en el Cuadro 4, mientras que *esquilmar* presenta 148 ocurrencias en el CORPES XXI, un sinónimo de esta palabra, como podría ser *arrui-*

*nar*, aparece 2149 veces. *Verbigracia* tiene 86 ocurrencias, mientras que su sinónimo *por excelencia*, por ejemplo, tiene 41 282. En lo que respecta a *arramplar*, se ha comparado el uso de esta palabra con la forma “preferible” según el DRAE, que es *arramblar*. Así, mientras que de la primera forma solo hay 8 coincidencias en el CORPES XXI, de la segunda se presentan 45 casos. La comparativa de este cuadro nos permite determinar que los términos que utiliza el autor de los textos indubitados son relativamente “poco habituales” en español y, por tanto, formas marcadas de su estilo idiolectal.

CUADRO 4. Frecuencia de aparición en CORPES XXI de los términos usados en los textos indubitados

<i>Esquilmar</i>	<i>Arruinar</i>	<i>Verbigracia</i>	<i>Por excelencia</i>	<i>Arramplar</i>	<i>Arramblar</i>
148	2149	86	41 282	8	45

Por lo que concierne a la imitación de estos términos (Cuadro 5), se puede concluir que los autores han tendido a no reproducirlos en la mayoría de los casos. El término *esquilmar* ha sido reproducido por tres autores, *verbigracia* no aparece en ninguna imitación y únicamente un autor ha recogido *arramplar*.

CUADRO 5. Ejemplos de palabras poco frecuentes

CORPUS INDUBITADO	CORPUS SIMULACIÓN
Políticas que esquilman	Han arramplado con... [11]
La esquilmación de Caja Navarra	Podrá ser esquilmada [12]
A Grecia, verbigracia, se le da dinero para cubrir...	Va usted a esquilmar sus posibilidades de sanar [15]
Hay que arramplar con la institución	Esquilmar a los que han otorgado la confianza [17]

### 7.1.3. Anglicismos

Una variable que no ha sido reproducida por ningún imitador es el uso de anglicismos o expresiones en inglés (por ejemplo, *shock*, *run on the bank* u *Occupy Wall Street*), hápax legómena<sup>6</sup> seguidas de una explicación del significado en español.

<sup>6</sup> Palabra que aparece una única vez en un corpus.

#### 7.1.4. Frases hechas

El autor de los textos indubitados utiliza frases hechas y, como ya se ha analizado en el léxico, también alterna registros formales e informales. Además, en alguna ocasión adapta las expresiones a su propio estilo, al igual que sucedía con su elección de vocablos (*arramplar* vs. *arramblar*), en este caso, modifica la expresión más habitual en español según el DRAE *tirarse a degüello* por *lanzarse a degüello*.

En el corpus de simulación encontramos que varios autores han reproducido metáforas del corpus indubitado e incluso han introducido alguna nueva, tal como se puede observar en el Cuadro 6. Es notable que la expresión que han reproducido en más ocasiones es *lanzarse a degüello* y que uno de los autores, a pesar de querer imitar su uso, ha empleado su propio estilo, puesto que la ha modificado nuevamente y ha escrito *lanzarse al degüello*.

CUADRO 6. Ejemplos de frases hechas

CORPUS INDUBITADO	CORPUS SIMULACIÓN
A la postre	Ver la luz [11]
Arrimar el hombro	Ser peor el remedio que la enfermedad [11]
Pasar a la historia	Arrimar el hombro [12]
Lanzarse a degüello	Lanzarse a degüello [13]
Morderse la lengua	Estar a la altura [17]
Estar a la altura	Meter la pata [17]
	Lanzarse al degüello [18]

#### 7.1.5. Expresiones

En este apartado se analizan expresiones prominentes del autor. Esta variable solo ha sido reproducida por uno de los autores en la imitación. En particular, el autor ha utilizado expresiones e incluso oraciones exactamente iguales a las usadas por el autor de los textos indubitados. En el Cuadro 7 se muestran algunos ejemplos.

CUADRO 7. Ejemplos de expresiones

CORPUS INDUBITADO	CORPUS SIMULACIÓN
El negocio bancario, bien llevado, es rentable	El negocio bancario bien llevado puede sanear [113]

CUADRO 7. Ejemplos de expresiones (*continuación*)

CORPUS INDUBITADO	CORPUS SIMULACIÓN
Pues bien, el dinero que ahora nos llega es dinero público, de nuestros socios europeos	Pues bien, el dinero que ahora está por llegar es dinero público de nuestros socios europeos [I13]
Lo vivido ayer es un momento histórico, de mayor profundidad de lo que nadie puede calibrar todavía, hoy que salimos de la resaca del evento	Lo que vivimos ahora es un momento histórico de una profundidad incalculable [I13]
Un país que se debate entre su capacidad de volver a ser grande o terminar deambulando en lo misérrimo	Ya que el país se debate entre volver a ser grande o bien terminar deambulando en lo misérrimo [I13]

## 7.2. Morfosintáctico

### 7.2.1. Orden de palabras: adjetivo antepuesto

El autor de los textos indubitados muestra una preferencia por la anteposición al nombre de la mayor parte de los adjetivos calificativos. Esta variable ha sido detectada por muy pocos imitadores y, en el caso de ser detectada, ha sido reproducida con una frecuencia mucho menor, por lo que no se puede descartar que haya sido una anteposición espontánea. Se muestran algunos ejemplos en el Cuadro 8.

CUADRO 8. Ejemplos de adjetivo antepuesto

CORPUS INDUBITADO	CORPUS SIMULACIÓN
legítima mayoría, poco original remedio, enfática referencia, deletéreos efectos, pasada legislatura, general conocimiento, presunto poder, actual Gobierno, mayor problema, próximos presupuestos, mejor sistema, mayor burbuja, mero hecho	posibles pactos [I1], próximos meses [I1], lacerantes cambios [I1], determinantes medidas v propia alternativa [I1], exitosos frutos [I1], sucesivos meses [I1], pasadas elecciones [I2], mayor número [I2], siempre oportunista PSN [I2], hasta ahora protegida y saneada Caja de Navarra [I2], antiguas mayorías [I2], mero hecho [A3], buena alumna [I4], mayor burbuja [I6], maldito entramado [I6], legítima mayoría [I6], mordaz pluma [I6]

## 7.2.2. Perífrasis de obligación

Todavía desde la perspectiva del análisis morfosintáctico, un aspecto lingüístico que presenta cierto polimorfismo —y que, por lo tanto, puede ser objeto de análisis en una comparación forense de textos— es la expresión de la obligación. La lengua española cuenta con múltiples perífrasis para indicar obligación. Las más usuales son *deber* + infinitivo, *tener que* + infinitivo, *haber que* + infinitivo y *haber de* + infinitivo.

El corpus indubitado destaca por el uso de tres perífrasis (*deber/haber que/haber de* + infinitivo) con un claro predominio de la primera, como se observa en el Cuadro 9. Ninguno de los imitadores ha utilizado las tres perífrasis de obligación: o se han decantado por la más destacada o han utilizado perífrasis inexistentes en los textos indubitados o incluso cometido errores como el uso de *deber* + infinitivo por *deber de* (que, según el *Diccionario panhispánico de dudas*, “denota probabilidad o suposición”).

CUADRO 9. Ejemplos de perífrasis de obligación

PERÍFRASIS DE OBLIGACIÓN	CORPUS INDUBITADO	CORPUS SIMULACIÓN
<i>Deber</i> + infinitivo	se les debe ayudar —en contadas ocasiones— ...se les deba ayudar las cosas deben ir la banca debe ser saneada debe contar con la disponibilidad debiera proscribirse el uso debiera gustarles el cambio	debería ser puesto de antemano en conocimiento de la fiscalía [11] alguien debe advertir del error [15] las cosas deben ir por otro lado [16]
<i>Haber que</i> + infinitivo	hay que deslegitimar al conjunto del sistema hay que arrampalar con la institución habrá que imputar el pago	—
<i>Haber de</i> + infinitivo	hemos de mejorar nuestro trabajo cada día	hemos de mejorar nuestro trabajo cada día. [15] el mundo ha de saber de su incompetencia [17]
<i>Tener que</i> + infinitivo	—	tendría que saber [14]
<i>Deber de</i> + infinitivo	—	deberían de producirse [17]

### 7.2.3. Imperfecto de subjuntivo

En español existen tiempos verbales que admiten más de una realización, y la elección de una forma u otra puede depender de factores dialectales o incluso idiolectales. Un ejemplo es el pretérito imperfecto de subjuntivo, que se puede materializar con los afijos *-ra-* o *-se-*. Los textos indubitados muestran una preferencia por *-ra-*, y esto ha sido una característica que, aun estando ausente en algunas de las imitaciones, se ha mantenido en la mayoría de los textos que contenían algún verbo en este tiempo.

CUADRO 10. Ejemplos de imperfecto de subjuntivo

CORPUS INDUBITADO	CORPUS SIMULACIÓN
No generaran	Pudieran haber escapado al control de este Gobierno [12]
Debiera proscribir el uso	Pudiera derivar o acontecer [12]
Debiera gustarles el cambio, ¿no?	Si hubiera sido una buena alumna [14]
Si de golpe todos ellos quisieran retirar su dinero	Lo honesto hubiera sido no silenciar esa gran mayoría... [17]

### 7.2.4. Expresión de futuro

Otro caso de variedad de realizaciones se encuentra en la expresión de futuro, que se puede materializar mediante una forma verbal del futuro o mediante la perífrasis verbal *ir + a +* infinitivo. Ningún imitador ha reproducido con exactitud esta característica y, de hecho, se inclinan mayoritariamente por las formas de futuro simple.

CUADRO 11. Ejemplos de expresión de futuro

EXPRESIÓN DE FUTURO	CORPUS INDUBITADO	CORPUS SIMULACIÓN
Forma verbal	Habrá que sacar	se irá [113]
	Podrá aumentar	tendrá [113]
	Habrá que imputar	cooperará [115]
	Las cosas podrán cambiar para siempre	te verás [120]
	Si será posible	estarán [122]
	Costará, pero se limpiará	harán [128]
		entenderé [128]

CUADRO 11. Ejemplos de expresión de futuro (*continuación*)

EXPRESIÓN DE FUTURO	CORPUS INDUBITADO	CORPUS SIMULACIÓN
Perífrasis verbal <i>ir + a +</i> infinitivo	Nos van a aportar Vamos a comprar Se va a poner orden Vas a tener	Va a conseguir [I29] Va a creer [I29] va a hacerlo [I30] va a depositar [I30]

### 7.2.5. Pronombres relativos

Como han mostrado trabajos anteriores, el análisis del uso de relativos puede ser una variable de peso en la comparación con fines de atribución de autoría y en la construcción de perfiles lingüísticos. Por ejemplo, en el análisis del uso de los relativos en textos en español producidos por bilingües catalán-español puede ser de interés analizar la distribución de los relativos *que* y *cual*, dado que la alta frecuencia del relativo compuesto con *cual* (*qual*) en catalán puede influenciar sobre el uso de dicho relativo en español (Turell, 2010). También puede ser significativa la elección por el relativo *quien* o *que* en cláusulas relativas de persona —lo que seguramente es una cuestión idiosincrásica, más que un fruto del contacto entre lenguas.

El análisis de los textos indubitados muestra un uso bastante equilibrado de *que* (43%) y *quien* (57%) en las relativas de personas, aunque con una ligera preferencia hacia *quien*. En el resto de relativas específicas el autor usa siempre *que*. Ninguno de los imitadores ha sabido reproducir esta variable. Concretamente, y como se observa en el Cuadro 12, solo dos imitadores han reproducido, en una única ocasión cada uno (es decir, en un 17 y 20%), *quien* e incluso ha habido imitadores que han incluido el uso de *el cual*.

CUADRO 12. Ejemplos de pronombre relativos

PRONOMBRE RELATIVO	CORPUS INDUBITADO	CORPUS SIMULACIÓN
<i>Quien</i>	Quienes actuamos Quienes han tomado Quienes más hablan Para quienes estos días	Quienes más hablan [I17] Quienes dijeron [I13]
<i>Que</i>	nadie de los que ayer que son menos de los que toman Neptuno	al lado de los que quieren destruir [I14] los que sabemos [I14] para los que hemos [I5]

CUADRO 12. Ejemplos de pronombre relativos (*continuación*)

PRONOMBRE RELATIVO	CORPUS INDUBITADO	CORPUS SIMULACIÓN
<i>Que</i>		cómplice de los que con su gestión [15] A los que han otorgado [17]
<i>Cual</i>	–	Izquierda totalitaria [...], la cual será incapaz [11] me refiero propiamente al cual su partido [11] ninguno de los cuales ha producido [11] momento en el cual [11] pacto [...] al cual pretende añadirse Caja de Navarra [12] La cual podrá [12] [...] lo cual debería [12]

### 7.3. Pragmático-discursivo

#### 7.3.1. Marcadores discursivos en posición inicial

Los textos indubitados se caracterizan por una baja presencia de marcadores discursivos y por una escasa variedad de los mismos. En concreto, para reforzar la argumentación el autor utiliza *pues bien*, *en efecto*, *sin duda* (siempre escritos entre comas o seguidos de coma). Se pueden ver algunos ejemplos en el Cuadro 13. Para expresar consecuencia, se emplean las locuciones *como consecuencia de* y *por consecuencia*. Algunos de los imitadores han prestado atención a qué marcadores utiliza el autor de los textos indubitados, sin embargo, en la mayoría de las ocasiones no han prestado atención a la posición o a la puntuación. En lo que concierne a la expresión de la consecuencia, ninguno ha reproducido la locución exacta, sino que se han decantado por fórmulas como *a consecuencia de* o *en consecuencia*.

CUADRO 13. Ejemplos de marcadores discursivos para reforzar la argumentación

CORPUS INDUBITADO	CORPUS SIMULACIÓN
Pues bien, el que ahora nos llega	(en efecto, me refiero...). [1]
En efecto, son menos que...	Ellos buscan, sin duda, la ocasión de... [19]
sin duda, molestas...	sin duda de nuevo... [10]
Buscan, sin duda, la ocasión	sin duda ahora [10] Pues bien, señora mía [15] Es aceptar, sin duda, la ocasión para... [15] Esta es sin duda su actual [27] sin duda pretende [27]

### 7.3.2. Expresión del énfasis

Otra variable a tener en cuenta desde el punto de vista pragmático-discursivo es cómo se marca tipográficamente el énfasis. Hay que señalar, en este sentido, que en los textos indubitados se recurre al uso de la negrita tal y como se ilustra en el Cuadro 14. La mayoría de los imitadores han sabido detectar y reproducir esta variable. No obstante, algunos de ellos han recurrido a otras formas de marcación del énfasis como el uso de la mayúscula o la repetición de palabras.

CUADRO 14. Ejemplos de expresión del énfasis mediante negrita

CORPUS INDUBITADO	CORPUS SIMULACIÓN
ocupar el Congreso	económicos y políticos [116]
largo tiempo de maceración	renuncie y disuelva su formación [117]
apenas 6.000 personas	taquilla número 201 [119]
intencionalidad política poco disimulable	Chamartin [122]
expresión de la izquierda totalitaria	martes 23 a las 12:00 h. [127]
Una quinta parte del tiempo que corresponde al mandato.	
Hay vallas que limitan el acceso viario y control policial.	
no es la primera vez que se adoptan estas medidas	
el blanco de sus protestas	

### 7.3.3. Formato de incisos

Como se muestra en el Cuadro 15, en los textos indubitados los incisos se expresan mediante el paréntesis y los guiones, que siempre son guiones cortos (–) en lugar de largos (—), uso recomendado por la Real Academia Española (Asale, 2005) y que, por lo tanto, sería lo menos marcado. Un número reducido de imitadores han reproducido ambos recursos, y todavía menos lo han hecho de forma exacta. La mayoría de los imitadores han realizado los incisos entre comas.

CUADRO 15. Ejemplos de formato de incisos

CORPUS INDUBITADO	CORPUS SIMULACIÓN
cuando desarrolla sus funciones —es delito trastornarlas violentamente— y tornaron la idea	será peor el remedio que la enfermedad —que en este caso la enfermedad parece haber calado hasta lo más hondo de la sociedad— al ver los resultados en las urnas electorales. [11]

CUADRO 15. Ejemplos de formato de incisos (*continuación*)

CORPUS INDUBITADO	CORPUS SIMULACIÓN
las cúpulas sindicales de UGT y CC.OO. —sin duda, molestas por la reducción de sus 37 liberados— convocan para mañana una manifestación	un sistema financiero y político como el actual (en efecto, me refiero propiamente al cual su partido quiere destruir). [1]
Se le debe ayudar —en contadas ocasiones— no tanto por el riesgo moral	que usted haya colocado su partido, —socialista, obrero y español recordemos—, al lado de los que quieren destruir la unidad de la patria [14]
manifestarse frente al Congreso (véanse sus soflamas y pasquines) no pretenden sólo que se escuche su voz. los años del trueque ya pasaron a la historia (aunque algunos no quieran enterarse).	

## 7.4. Complejidad

### 7.4.1. Media de palabras

Dos rasgos que han mostrado ser idiosincrásicos son la longitud de frase y de párrafo en términos de palabras por unidad. Ningún imitador ha reproducido con exactitud el número medio de palabras por frase, aunque unos pocos se han aproximado. En la Figura 2 se muestra los resultados de la media de palabras por frase. Los valores del autor indubitado aparecen en última posición y se ha

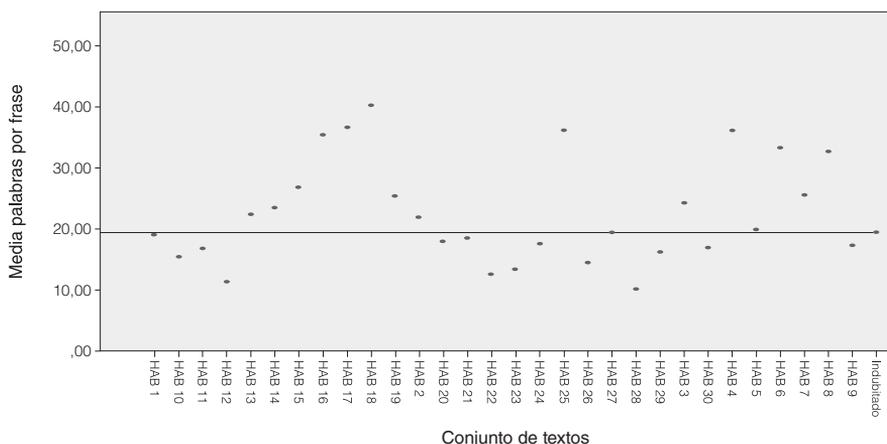


FIGURA 2. Media de palabras por frase en textos de imitación e indubitado

trazado una línea con su valor para poder visualizar de forma más sencilla qué autores son más capaces de reproducir el rasgo.

Se procede a realizar la prueba de contraste  $t$  con un valor de prueba 19.35, correspondiente al valor medio de los textos indubitados, que es de 19.35 palabras por frase. En el Cuadro 16 se observa que el valor del estadístico  $t$  de la prueba es de 2.238 con un  $p$ -valor = 0.033, por lo que se puede concluir que la media de palabras por frase es significativamente distinta. Además, en el cuadro también tenemos la diferencia de medias (diferencia entre la media muestral de palabras por frase y el valor de referencia 19.35) y el intervalo de confianza para esta diferencia.

CUADRO 16. Resultados estadísticos media de palabras por frase

	$T$	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Media de palabras por frase	2.238	29	0.033	3.47667	0.2989	6.6544

Nota: Valor de prueba = 19.35

95% de intervalo de confianza de la diferencia

## 8. Resultados

Los resultados de las variables analizadas en este estudio permiten concluir que muy pocos rasgos han sido reproducidos por más de la mitad de los imitadores, específicamente, solo tres rasgos. Los rasgos más fáciles de reproducir han sido el léxico especializado, la expresión del énfasis y el imperfecto de subjuntivo con porcentajes de éxito en la imitación de 86.7, 83.3 y 66.7, respectivamente. En concreto, 26 de los 30 imitadores han sido capaces de identificar el uso de léxico especializado y lo han integrado en su texto, 25 han expresado el énfasis mediante negrita al igual que el autor de los textos dubitados y 20 de los 30 han utilizado el imperfecto de subjuntivo con la terminación *-ra*.

Por el contrario, los rasgos imposibles de emular para los imitadores de este estudio han sido principalmente las variables morfosintácticas (los pronombres relativos, la perifrasis de obligación, el futuro) y léxicas (los anglicismos). En el caso de las cuatro primeras variables anteriores ningún imitador ha sido capaz de reproducirlas. Con un bajo porcentaje de éxito encontramos la variable de

complejidad, solo 5 imitadores han sido capaces de reproducir la media de palabras por frase; en el caso de las variables pragmático-discursivas, solo dos imitadores han reproducido los marcadores discursivos con éxito y 5, el formato de incisos; en cuanto a las variables léxico-semánticas, solo un imitador reprodujo el uso de expresiones, y, por último en el caso de las variables morfosintácticas, 4 participantes imitaron exitosamente el uso del adjetivo antepuesto. En el Cuadro 17 se muestran los porcentajes de éxito en la imitación de cada una de las variables analizadas en este estudio ordenadas de mayor a menor éxito.

CUADRO 17. Porcentaje de éxito en la imitación de las variables

VARIABLES	PORCENTAJE DE ÉXITO EN LA IMITACIÓN
Léxico especializado	86.7
Expresión del énfasis	83.3
Imperfecto de subjuntivo	66.7
Frasas hechas	40.0
Palabras poco frecuentes	16.7
Media de palabras por frase	16.7
Formato incisos	16.7
Orden de palabras: adjetivo antepuesto	13.3
Marcadores discursivos en posición inicial	6.7
Expresiones	3.3
Anglicismos	0.0
Futuro	0.0
Perífrasis de obligación	0.0
Pronombres relativos	0.0

Como se ha podido comprobar en algunos de los ejemplos, en la mayoría de los casos el imitador deja rastro de su propio estilo. Algunos de los rastros detectados han sido otras variantes de la misma variable como, por ejemplo, la expresión del énfasis mediante mayúscula, incisos entre comas u otras formas de perífrasis verbales. También se han detectado otras variables relacionadas con los errores presentes únicamente en algunos de los textos del corpus de simulación. El uso de estructuras lingüísticas y de ortografía que se adhieren o se desvían de las convenciones y de la normativa (lo que desde un punto de vista

normativo se denominan usos correctos y erróneos, respectivamente) que utiliza un individuo puede dar pistas muy importantes sobre cuestiones sociolingüísticas relacionadas con su nivel educativo o ser resultado de fenómenos de contacto de lenguas o de distintos grados de familiaridad del individuo con el tipo de texto o género textual que intenta reproducir. Algunos de los errores detectados en el corpus de simulación y que nos indicarían que no han sido producidos por el mismo autor en el caso de una pericial de comparación forense de textos escritos serían: la no acentuación de pronombres interrogativos, el uso de la conjunción *i* en lugar de *y* (seguramente provocado por la influencia del catalán en el estilo idiolectal del imitador), uso de coma entre sujeto y verbo o el uso del punto después del signo de interrogación, entre otros.

## 9. Discusión

Los resultados obtenidos en este estudio muestran que los rasgos con un mayor éxito en la imitación son variables visualmente prominentes, aunque pertenecen a tres niveles diferentes (léxico, pragmático-discursivo y morfosintáctico). Solo la tercera variable más fácilmente reproducida, el imperfecto de subjuntivo, podría considerarse más directamente relacionada con los conocimientos de lengua de los participantes. Aun así, sería interesante explorar si variables como el uso de negrita o el uso de léxico especializado de medicina en un texto sobre política podrían ser detectadas en un nivel comparable a los resultados expuestos aquí por hablantes sin una formación universitaria en disciplinas relacionadas con la lingüística. En el futuro, y siguiendo la línea de investigación propuesta aquí, es probable que una comparación del rendimiento de grupos de participantes distintos, según su formación, pueda aportar más información sobre el papel que esta juega en la detección e imitación de variables lingüísticas.

Por otro lado, se ha comprobado que las variables morfosintácticas, así como ciertas variables léxicas, no han sido imitadas por los participantes. Cabe destacar que la dificultad de detección demostrada aquí de dichas variables apoya al objetivo de este estudio, ya que conforman ejemplos típicos de las variables observadas por los lingüistas forenses durante la comparación de textos escritos. Asimismo, el bajo porcentaje de éxito de las variables de complejidad, marcadores discursivos, posición de los adjetivos y usos de puntuación (como en los in-

cisos) también confirma la hipótesis de partida de los peritos lingüistas, ya que, mientras unos pocos autores en este estudio han conseguido aproximarse a los usos del imputado, ninguno de ellos ha conseguido imitar todas las variables con un porcentaje que le permitiera suplantar al sujeto sin ser detectado por un análisis sistemático y meticulado.

Así pues, las observaciones que se desprenden de los resultados obtenidos corroboran la validez de las hipótesis de trabajo, ya que, aun sin limitaciones de tiempo, los participantes no consiguieron emular satisfactoriamente el estilo idiosincrásico del corpus indubitado, lo cual indica que, efectivamente, el proceso de imitación requiere, entre otros, de un gran dominio lingüístico, capacidad de atención a la forma y una lectura consciente. Como se ha demostrado, para un análisis forense de comparación de textos escritos que permita la detección y caracterización de los autores es indispensable el estudio de múltiples niveles lingüísticos y variables. Finalmente, también se ha observado que ciertos autores han impregnado con su propio estilo idiolectal las variables que detectaron en el corpus indubitado (por ejemplo, perífrasis verbales).

En lo que se refiere a los métodos empleados en este estudio, a pesar de las limitaciones descritas, derivadas, por ejemplo, de la carencia de un corpus de referencia adecuado para los propósitos del estudio, para algunas de las variables se ha podido conseguir un cálculo (al menos orientativo) de la marcaje y prominencia de los usos lingüísticos. Estos cálculos permiten evaluar la probabilidad de que la reproducción de las variables en las producciones de los participantes sean fruto de la detección de características lingüísticas del estilo idiolectal del imputado o de su propia individualidad lingüística. Para las demás variables, el análisis cualitativo y el uso de la estadística descriptiva también han permitido observar la prominencia de las variables y su índice de imitación por parte de los participantes. Este tipo de análisis continua siendo, pues, indispensable mientras el uso de corpus de referencia adecuados y la obtención de la distribución poblacional no sea posible para todas las variables observadas.

## 10. Conclusiones

Este trabajo ha aportado datos empíricos que demuestran las dificultades de imitar el estilo idiolectal. Este estilo tan característico de cada individuo, único

e irreplicable, no se compone únicamente de una variable, sino que es precisamente todo el conjunto de variables analizadas lo que caracteriza el uso del lenguaje particular y único de un autor determinado. De este modo, se ha podido comprobar que la naturaleza de un discurso se configura por toda una serie de rasgos cuya concurrencia en un texto pasa inadvertida a cualquier usuario de la lengua hablada y es altamente improbable que se pueda imitar con la precisión suficiente para no ser detectado en un riguroso análisis comparativo de textos, ya sea mediante métodos cualitativos o una combinación de estos con métodos cuantitativos.

Los resultados de este estudio permiten concluir que existe una alta dificultad para emular con precisión el estilo característico de otra persona. Como se ha señalado al inicio, el estudio de la bibliografía disponible apunta a la necesidad de colaboración interdisciplinaria que nos ayude a comprender mejor los procesos cognitivos involucrados en la imitación de textos producidos por otros autores. Aun así, los resultados de este estudio muestran que los imitadores han sido capaces de reproducir algunos de los rasgos más visibles, pero difícilmente han podido aislarlos e incorporarlos todos y, mucho menos, han podido utilizarlos con la misma frecuencia.

Además, la mayoría de los imitadores no ha podido evitar dejar el rastro de su propio estilo idiolectal. Así pues, se ha podido comprobar que el estilo idiolectal de un individuo puede resultar tan inherente que se refleja en cada discurso que realiza.

En conclusión, este estudio apunta hacia la posibilidad de refutar el recurrente argumento a favor de la posibilidad de imitación del estilo del autor. Se ha demostrado que dicha suplantación resultó difícil de realizar a todos los participantes aun sin limitaciones de tiempo y mediante el uso de un corpus indubitado al que probablemente recurrirían en caso de querer suplantar su identidad. Además, se ha confirmado que la imitación es detectable para un lingüista forense mediante el análisis de las variables adecuadas y la implementación de la metodología protocolaria.

## 11. Referencias

- ABECASSIS, MICHAEL (2002). Saliency and frequency in a corpus of 1930's French films. *California Linguistic Notes*, 27(2), 1–18.
- ADANK, PATTI; HAGOORT, PETER, & BEKKERING, HAROLD (2010). Imitation improves language comprehension. *Psychological Science*, 21(12), 1903–1909.
- ARRIAGA RAMÍREZ, J. C. PEDRO; ORTEGA SAAVEDRA, MARÍA GUADALUPE; MEZA REYNOSO, GUSTAVO; HUICHÁN OLIVARES, FRANCISCO; JUÁREZ MALDONADO, EDITH; RODRÍGUEZ CUADROS, ADRIANA, & CRUZ MORALES, SARA E. (2006). Análisis conceptual del aprendizaje observacional y la imitación. *Revista Latinoamericana de Psicología*, 38(1), 87–102.
- ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (Asale) (2005). *Diccionario panhispánico de dudas*. Recuperado de <http://lema.rae.es/dpd>
- BERNÁRDEZ, ENRIQUE (2003). El texto en el proceso comunicativo. *Revista de Investigación Lingüística*, 6(2), 7–28.
- BLASS, REGINA (1990). *Relevance relations in discourse: A study with special reference to Sissala* (Vol. 55). Cambridge: Cambridge University Press.
- BORENSTEIN, ELHANAN, & RUPPIN, EYTAN (2005). The evolution of imitation and mirror neurons in adaptive agents. *Cognitive Systems Research*, 6(3), 229–242.
- BRUNNER, ELIZABETH GENTRY (2011). *Imitation, awareness, and folk linguistic artifacts* (Tesis doctoral inédita). Rice University, Houston.
- BUTLER, CHRIS (1998). Enriching the Functional Grammar lexicon. En Hella Olbertz, Kees Hangeveld & Jesús Sánchez García (Eds.), *The structure of the lexicon in functional grammar* (pp. 171–194). Ámsterdam: John Benjamins.
- CANTER, DAVID, & CHESTER, JOANNE (1997). Investigation into the claim of weighted Cusum in authorship attribution studies. *Forensic Linguistics. The International Journal of Speech, Language and the Law*, 4(2), 252–261.
- CHASKI, CAROLE E. (2001). Empirical evaluations of language-based author identification techniques. *The International Journal of Speech Language and the Law*, 8(1), 1–65.
- COULTHARD, MALCOLM (1994). On the use of corpora in the analysis of forensic texts. *International Journal of Speech Language and the Law*, 1(1), 27–43.
- COULTHARD, MALCOLM (2004). Author identification, idiolect and linguistic uniqueness. *Applied Linguistics*, 25(4), 431–477.
- COULTHARD, MALCOLM, & JOHNSON, ALISON (Eds.) (2007). *An introduction to Forensic Linguistics: Language in Evidence*. Nueva York: Routledge.

- COUPLAND, NIKOLAS (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.
- EL BOUANANI EL MANAR, SARA, & KASSOU, ISMAIL (2014). Authorship analysis studies: A survey. *International Journal of Computer Applications*, 86(12), 22–29.
- FOSTER, BILL R., JR. (marzo, 1989). Classical imitation and reading/writing connections: Analysis and genesis enter the twentieth century. Ponencia presentada en la 40th Annual Meeting of the Conference on College Composition and Communication. Seattle, Washington.
- GEIST, UWE (2005). Stylistic imitation as a tool in writing pedagogy. En Gert Rijlaarsdam, Huub van der Bergh & Michel Couzijn (Eds.), *Effective learning and teaching of writing: A handbook of writing in education* (pp. 169–179). Boston: Kluwer.
- GRANT, TIM (2007). Quantifying evidence in forensic authorship analysis. *International Journal of Speech, Language & the Law*, 14(1), 1–25.
- GRANT, TIM (2010). Text messaging forensics. Txt 4n6: Idiolect free authorship analysis? En Malcolm Coulthard & Alison Johnson (Eds.), *The Routledge Handbook of Forensic Linguistics* (pp. 508–522). Nueva York: Routledge.
- HERNÁNDEZ ESTEBAN, MIREIA (2016). *Lingüística forense básica*. Valencia: Tirant lo Blanch.
- HEYES, CECILIA M. (1993). Imitation, culture and cognition. *Animal Behaviour*, 46(5), 999–1010.
- HOWALD, BLAKE STEPHEN (2008). Authorship attribution under the rules of evidence: Empirical approaches in a layperson's legal system. *The International Journal of Forensic Linguistics*, 15(2), 219–247.
- IACOBONI, MARCO (2009). Neurobiology of imitation. *Current Opinion in Neurobiology*, 19(6), 661–665.
- IRIZARRY, ESTELLE (1989). Exploring conscious imitation of style with ready-made software. *Computers and the Humanities*, 23(3), 227–233.
- JAKOBSON, ROMAN (1971 [1956]). Two aspects of language and two types of aphasic disturbances. En Roman Jakobson (Ed.), *Selected writings II: Word and Language* (pp. 240–259). La Haya: Mouton.
- JOHNSTONE, BARBARA (2000). The individual voice in language. *Annual Review of Anthropology*, 29, 405–424.
- McMENAMIN, GERALD (2001). Style markers in authorship studies. *The International Journal of Speech, Language and the Law*, 8(2), 93–97.
- MORTON, ANDREW Q. (1991). *Proper words in proper places*. Technical Report 911R18. Glasgow: University of Glasgow, Computing Science Department.

- OAKMAN, ROBERT LEE (1984). *Computer methods for literary research*. Athens: University of Georgia.
- QUERALT, SHEILA ESTEVEZ (2014). Acerca de la prueba lingüística en atribución de autoría hoy. *Revista de Llengua i Dret*, 62, 35–48.
- REAL ACADEMIA ESPAÑOLA (RAE). *Diccionario de la lengua española* (22ª ed.). Recuperado de <http://www.rae.es/rae.html>
- REAL ACADEMIA ESPAÑOLA (RAE), Banco de datos (CORPES XXI). *Corpus del español del siglo XXI (CORPES)*. Recuperado de <http://www.rae.es>
- REAL ACADEMIA ESPAÑOLA (RAE), Banco de datos (CREA). *Corpus de referencia del español actual*. Recuperado de <http://www.rae.es>
- SANDFORD, ANTHONY J.; AKED, JOY P.; MOXEY, LINDA M., & MULLIN, JAMES (1994). A critical examination of assumptions underlying the cusum technique of forensic linguistics. *Forensic Linguistics*, 1(2), 151–167.
- SOMERS, HAROLD (s. f.). *Stylometry and authorship* [diapositivas de PowerPoint]. Recuperado de <http://personalpages.manchester.ac.uk/staff/harold.somers/LELA30922/>
- STAMATATOS, EFSTATHIOS (2009). A survey of modern authorship attribution methods. *Journal of the Association for Information Science and Technology*, 60(3), 538–556.
- TRIOLA, MARIO (2004). *Probabilidad y estadística*. México: Pearson educación.
- TURELL, MARIA TERESA (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech, Language and the Law*, 17(2), 211–250.
- TURELL, MARIA TERESA (2011). La tasca del lingüista detectiu en casos de detecció de plagi i determinació d'autoria de textos escrits. *Llengua, Societat i Comunicació*, 9, 69–85.
- WILSON, DEIRDRE, & SPERBER, DAN (1993). Linguistic form and relevance. *Lingua*, 90(1–2), 1–25.
- WRIGHT, DAVID (2014). *Stylistics versus Statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails* (Tesis doctoral inédita). University of Leeds, Reino Unido. Recuperado de <http://theses.whiterose.ac.uk/8278/1/David%20Wright%20%28200327943%29%20Final%20Thesis%20-%20Stylistics%20versus%20Statistics%20%284.3.15%29.pdf>
- ZHENG, RONG; QIN, YI; HUANG, ZAN, & CHEN, HSINCHUN (2003). Authorship analysis in cyber-crime investigation. En Hsinchun Chen, Therani Madhusudan, Richard Miranda, Daniel D. Zeng, Jenny Schroeder & Chris Demchak (Eds.), *Proceedings of the 1st NSF/NIJ Conference on Intelligence and Security Informatics* (pp. 59–73). Berlín: Springer.