

LA IMITACIÓN COMO CONTRAARGUMENTO EN PERITAJES DE ATRIBUCIÓN DE AUTORÍA: ESTUDIO DE UN CASO

Resumen: En los últimos años han aumentado los informes periciales en lingüística forense presentados en vistas judiciales para asesorar al juez sobre la autoría de varios escritos. Las partes contrarias intentan a menudo poner en duda las conclusiones del peritaje lingüístico utilizando como argumento principal la posibilidad de que existiera un segundo autor capaz de reproducir el estilo del acusado. Este artículo aporta fundamentos científicos para refutar la hipótesis sobre la posibilidad de la imitación del estilo de un autor. Se investiga un caso real de chantaje a un cargo público en el que participó una de las autoras. A partir de los resultados, se evalúa la posibilidad de imitación y cuáles son los rasgos del corpus indubitado más fáciles y difíciles de imitar.

Palabras clave: contraargumentación; prueba evidencial; peritaje lingüístico; lingüística forense; semejanza

DECEPTION AS COUNTERARGUMENT IN AUTHORSHIP ATTRIBUTION REPORTS: A CASE STUDY

Abstract: In recent years, the number of forensic linguistics expert reports submitted in court hearings to assess judges about the authorship of written texts has increased. Opposing sides frequently try to call the linguistic report's conclusions into question using as the main argument the possibility that there could be a second author who was able to replicate the defendant's style. This article provides scientific foundations on which to rebut the hypothesis of the possibility of deception of an author's style. A real case of blackmail to a public official, in which one of the authors was involved, is observed. On the basis of the results, the possibility of deception is evaluated, as well as the features of the undisputed corpus which are easiest and most difficult to replicate.

Key words: counterargumentation; evidence; linguistic expert report; forensic linguistics; similarity

1 INTRODUCCIÓN

El auge en el uso de plataformas electrónicas para comunicarse, como *Whatsapp*, *Facebook*, *Twitter*, *Instagram*, y el incremento de softwares para anonimizar y dificultar el rastreo de identidades por internet se está 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 incrementado el número de periciales lingüísticas, ya que estas últimas no analizan la forma manuscrita del mensaje ni tampoco rastrean el origen del mensaje, 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 analizan 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 que existiera 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, en base a los principios teóricos utilizados en lingüística forense, así como al saber común. El objetivo de este estudio es servir como primer paso 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 CASO

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 cargo público en el que se formuló esta hipótesis, en la cual se basó gran parte de la defensa.

El caso sucedió en el año 2012 en la ciudad de Pamplona (Navarra) entre el exdiputado del PP, Santiago Cervera, y el expresidente de Caja Navarra, José Antonio Asiáin. El caso tuvo repercusión mediática y tanto detalles de los hechos como los textos analizados se hicieron públicos en varios medios de comunicación. En la Figura 1 se presenta la sucesión de los hechos:

29 de noviembre

- El Sr. Cervera recibe en su cuenta personal de correo electrónico un anónimo de un supuesto ex empleado de Caja Navarra donde se le ofrece "información absolutamente importante sobre el caso de Caja Navarra. (...) He conseguido documentos que pueden hacer que actúe la justicia" y se le facilitan instrucciones para recoger el sobre en la muralla de Pamplona. Véase Figura 2.

30 de noviembre

- El exdiputado responde y afirma que no acudirá porque "el sistema de recogida de la información que me ofrece es inaceptable". Aun así, dice: "En todo caso, estoy seguro de que usted podrá encontrar maneras de hacerme llegar esa documentación por vía directa e igualmente discreta". Véase Figura 2.

5 de diciembre

- El Sr. Asiáin abre su correo electrónico profesional y descubre que el 4 de diciembre había recibido un anónimo en el que se le exigían 25.000 euros para silenciar una información comprometedor sobre su antiguo despacho de abogados, según la cual saldrían a la luz supuestos trabajos facturados para la entidad financiera que ahora preside de las que se habría beneficiado su hijo Diego, que trabaja en su mismo bufete. La fecha límite para la entrega sería el viernes 7 de diciembre, a las 22.00 horas. El presidente de Caja Navarra decide denunciar los hechos a la Guardia Civil de Morro Jable (Fuerteventura), donde estaba de vacaciones. Desde allí, se comunica a la comandancia de Pamplona, que pone en marcha la operación. Véase Figura 3.

9 de diciembre

- El Sr. Cervera acude al lugar señalado por el mensaje anónimo, junto a la muralla de Pamplona. Pasa dos veces por la zona, la segunda vez se cubre el rostro "con gorro y bufanda" al percatarse de que había una cámara de seguridad. Según el exdiputado, actuó de ese modo porque pensaba que la información podría tener "relevancia judicial". La Guardia Civil le estaba esperando y fue detenido. Seguidamente, compareció ante el juez y fue puesto en libertad sin haber prestado declaración formal puesto que todavía era aforado por su condición de diputado.

FIGURA 1. Sucesión de los hechos del caso



FIGURA 2. Correo electrónico anónimo y respuesta de Santiago Cervera

Fuente: http://politica.elpais.com/politica/2012/12/15/actualidad/1355603805_579701.html

Señor José Antonio Asiain.

Tengo en mi poder diversa documentación que demuestra que usted realiza para Caja Navarra determinadas actividades profesionales a través de su despacho de abogado. Entre esas actividades hay un buen número de casos de desahucios hipotecarios. Usted lleva tiempo facturando considerables cantidades de dinero a Caja Navarra a pesar de ser su presidente y prestando servicios de los que también participa y se lucra su hijo. El conocimiento público de las pruebas de que dispongo le causaría un serio problema por la sensibilidad social que existe hacia los abusos de las entidades financieras, el lucro de sus dirigentes y especialmente por todo lo relativo a los desahucios hipotecarios. La manera de evitar que se difunda la información y los documentos de que dispongo es la entrega de 25.000 euros. El lugar de la entrega se encuentra indicado al final de esta comunicación. La cantidad debe estar compuesta por billetes de alta denominación y contenida en un sobre oscuro convenientemente compactado y cerrado

mediante plástico para que resulte impermeable. El sobre debe depositarse antes de las 22 horas del próximo día 7 en el lugar que le explico.

Imagino que esta circunstancia le desagrade pero espero que la entienda con la misma seriedad con que yo lo hago. La difusión de los documentos que tengo en mi poder haría que usted obligatoriamente tuviera que renunciar a la presidencia de Caja Navarra y perjudicaría mucho a su amigo el señor Goñi, a otras varias personas de la antigua Banca Cívica y al partido político del que usted procede. La cantidad que le solicito no es negociable y sabe que es pequeña en relación con lo que usted viene facturando y con lo que pretende seguir facturando en el futuro. Puede considerarla como un pago imprescindible para mantener su situación actual y la de otras personas con las que ha compartido y comparte intereses.

Debo indicarle que una vez recibida la cantidad y superados unos días de margen de seguridad procederé a destruir los documentos, usted no volverá a tener noticias más y podrá seguir como hasta ahora. También le indico que si no cumple lo que le solicito, denuncia el caso o intenta vigilar o interferir en cualquier modo la entrega, la documentación será difundida automáticamente a diversos destinatarios entre los que están la organización Kontuzl, varios medios de comunicación locales y estatales y diversos cargos políticos. Los documentos están alojados en un servidor programado para hacer el envío salvo que reciba instrucción en contrario.

La reitero que este asunto se puede resolver fácilmente, pero todo lo que usted quiera complicarlo será perjudicial tanto para sus intereses como para los de otras personas. Tiene una única oportunidad. No mantendré ninguna nueva comunicación con usted. Lugar de la entrega. Se toma el ascensor público que se encuentra frente a la entrada del Club Natación y se sube hasta el fortín de san Bartolomé. Al salir del ascensor se cruza por una pasarela de unos veinte metros. A la izquierda se ve el muro del fortín que mira hacia Txantrea. En ese mismo muro hay dos rendijas verticales prácticamente a ras de la hierba. La que está a la derecha es más profunda y frente a ella hay un foco de suelo. En esa rendija se debe introducir el sobre con cuidado de que quede suficientemente disimulado y no pueda verse desde el exterior, y se debe tapar la rendija con hojas o hierba.

FIGURA 3. Correo electrónico anónimo recibido por el Sr. Aisaín

Fuente: http://www.navarraconfidencial.com/wp-content/uploads/2014/03/auto_cervera.pdf

El Sr. Aisaín solicitó un peritaje lingüístico de atribución de autoría en el que se debía comparar el texto anónimo con un corpus de textos del principal sospechoso, Santiago Cervera.

Durante el proceso de instrucción, el Magistrado-Juez instructor recabó indicios que sugerían que la persona que redactó y envió dicho mensaje pudiera ser el imputado del caso, esto es,

el Sr. Santiago Cervera Soto. Así pues, uno de los aspectos a considerar en el juicio era si el texto en cuestión podría atribuirse o no al imputado, Sr. Cervera.

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 ESTUDIO

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

a) Pregunta del letrado:

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:

¿Es posible que una tercera persona que conociera el estilo del Sr. Cervera imitara ese estilo para hacer creer que era el Sr. Cervera 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?

3.1 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. 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.

3.2 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 la 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 éstos queden impregnados inevitablemente de su propio estilo personal.

4 FUNDAMENTOS TEÓRICOS

El principio en el que se apoya este trabajo es el principio de “individualidad lingüística” (Johnstone, 2000; Chaski, 2001; Coulthard, 2004; Coupland, 2007; Howald, 2009; McMEnamin, 2001: 93–97; Grant, 2010) el cual postula que cada individuo se expresa mediante un “estilo idiolectal” (Turell, 2010) característico formado por la conjunción de sus características sociolingüísticas. La asunción de la hipótesis del “estilo idiolectal” 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 fundamental 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 consciente e 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 discurso. Una forma marcada será por tanto considerada una característica idiolectal, ya que es un elemento lingüístico no esperado o poco frecuente que se puede dar, por ejemplo, por contacto de lenguas. Un ejemplo de forma lingüística marcada sería el uso de un elemento léxico concreto en una posición en la que, según el comportamiento y/o características de su autor, se esperaría uno distinto.

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 en base a 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 muy 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 (ver p. ej. 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 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 ejemplariza el riesgo de delegar completamente en ecuaciones o softwares para la resolución de pruebas de autoría por suponer que su automaticidad y capacidad de ser generalizables garantiza su eficacia y fiabilidad. Además, la lectura de los resultados y gráficos proporcionados por este tipo de softwares 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). Aun así, en lingüística forense, es habitual contar con muestras lingüísticas pertenecientes a géneros diferentes (correos electrónicos en el trabajo, cartas de extorsión, publicaciones familiares en redes sociales, etc.) y, además, que los corpus comparados sean 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, 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.

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

5 MATERIAL DE ESTUDIO

Para llevar a cabo este trabajo se ha pedido a treinta personas con perfiles de edad y grupo social parecidos que redactaran una carta de amenazas intentando reproducir “el estilo” de Santiago Cervera, del que se proporcionaban cuatro textos extraídos de su blog (de ahora en adelante, corpus indubitado), con un total de 1.836 palabras. Estos textos y los que debían producir los participantes, pues, pertenecen a géneros diferentes. Sin embargo, como suele pasar en casos de atribución de autoría, no fue posible obtener producciones de Santiago Cervera más fácilmente comparables con las cartas de 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 casos de comparación forense de textos escritos para determinar si existe o no caso lingüístico (Queralt, 2014: 37).

5.1 *Corpus indubitado*

Los textos indubitados (TI) son cuatro textos extraídos del blog de Santiago Cervera (Cervera, 2012) 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
Lo del #25S	389
Congreso protegido.	341
Seis reflexiones sobre lo de ayer.	821
Tres reflexiones sobre rescates bancarios.	285

Se trata de un blog presumiblemente escrito personalmente por el Sr. Santiago Cervera, y que, en todo caso, sería la única referencia disponible en internet de textos presuntamente escritos por él. Por lo tanto, cabe pensar que, siendo ésta la única fuente directamente accesible, una persona que pretendiera simular el estilo de su autor, acudiría muy probablemente a ella como medio de inspiración.

5.2 *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 un tamaño de muestra mínimo específico que sea suficiente para todos los casos (Triola, 2004: 319) y, en este caso, 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 de 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.

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 delito. Sin embargo, el número de variables se debía acotar y se han analizado algunas de las variables que más se suelen utilizar, independientemente del contexto de estudio (*cf.* Turell, 2011; Queralt, 2014; Hernández Esteban, 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, los anglicismos, el uso de frases hechas y metáforas, así como las estructuras.

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 los incisos.

Como variables de **complejidad** se analiza el número de palabra por frase.

CUADRO 2. Variables de estudio

MÓDULO	VARIABLES
Léxico-semántico	Léxico especializado
	Palabras poco frecuentes
	Anglicismos
	Frases hechas
	Metáforas
	Expresiones
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 incisos
Complejidad	Media de palabras por frase

7 ANÁLISIS

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. Seguidamente, se ha realizado una comparación de esos rasgos con el corpus de textos de simulación. 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, creado por David Woolls de CFL Software.

Con el fin de valorar hasta qué punto 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 como: “¿cuántas realizaciones de la variable “adjetivo antepuesto” se observa 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 otros, 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 CREA o CORPES) no proporciona los datos requeridos para calcular la distribución poblacional o frecuencia esperable de dichas variables.

No obstante, en la mayoría de variables, nuestro objetivo era solamente observar si su índice de imitación en el corpus de simulación reproducía el uso identificado en el corpus indubitado, por lo que no fue necesario evaluar su marcadez y prominencia mediante medidas estadísticas o computacionales.

7.1 Léxico-semántico

El autor de los textos indubitados realiza un uso prominente de cuatro variables léxico-semánticas: el uso de vocabulario y metáforas del campo de la medicina, el uso de frases hechas y anglicismos. También destaca por el empleo de palabras poco habituales en español y, por tanto, marcadas.

7.1.1 Léxico especializado

El autor del conjunto de los textos indubitados recurre al léxico propio (y a metáforas propias) de 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
Bisturí	Me produce el mismo sentimiento que el de un médico que es consciente de que será peor el remedio que la enfermedad –
Ha salido pus, el empiema estaba ahí aunque algunos lo quisieran negar. Toca cirugía y no analgesia.	que en este caso la enfermedad parece haber calado hasta lo más hondo de la sociedad-
Un sistema, el financiero, aquejado de la gangrena de la corrupción.	No dispondré de calmantes para su dolor.
En contra de sus pulsiones.	Se producirá una hemorragia en el sistema financiero.
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.	En tus manos está pues que la reunión de ayer no tenga que generar unos efectos quirúrgicos irreversibles para tus intereses
No se puede poner al paciente a hacer ejercicio si no tiene sangre en las arterias. Se haya decidido a sajar con mayor profundidad del	la gangrena no se cura con analgésicos los médicos sabemos que la gangrena no se cura con analgésicos sino con una operación quirúrgica
	hay que aislar y extirpar la gangrena
	sustenta i apoya la gangrena [sic]
	parecía que nuestro país recuperaba las constantes vitales
	de nuevo se producirá una hemorragia en el sistema
	vacunarse frente a posibles represalias.

sistema financiero	provocar una amnesia permanente. sistema financiero óptimo,
disponibilidad de crédito	saqueo de la lógica financiera
inflábamos la mayor burbuja inmobiliaria	subvenciones públicas
reformas financieras	inversiones
presupuesto público	estimular la economía.
ajustes en su dispendio	inflaron los honorarios
pago de intereses	institución financiera gestiones ilícitas
políticas de estímulo económico	facturación de Caja Navarra
créditos suficientes para operar	malversación de fondos
actividad bancaria	expolio de la entidad actividad ilícita dentro de la entidad.

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 analizado mediante la lingüística de corpus analizando los términos usados por el autor de los textos indubitados y algunos sinónimos en el Corpus del Español del Sigo XXI (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.

CUADRO 4. Comparación frecuencia de aparición en CORPES XXI

ESQUILMAR	ARRUINAR	VERBIGRACIA	POR EJEMPLO	ARRAMPLAR	ARRAMBLAR
148	2149	86	14 282	8	45

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 *arruinar* 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 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 del estilo idiolectal del autor.

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 esquilmaban	Han arramplado con...
La esquilmación de Caja Navarra	Podrá ser esquilmada
A Grecia, verbigracia, se le da dinero para cubrir...	Va usted a esquilmar sus posibilidades de sanar
Hay que arramplar con la institución	Esquilmar a los que han otorgado la confianza

7.1.3 Anglicismos

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

7.1.4 Frases hechas

El autor de los textos indubitados también utiliza frases hechas y, como ya se ha analizado en el léxico, también alterna registros formales e informales. Además, en alguna ocasión también 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 ha utilizado 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	Arrimar el hombro
Arrimar el hombro	Lanzarse a degüello
Pasar a la historia	Ver la luz
Lanzarse a degüello	Ser peor el remedio que la enfermedad
Morderse la lengua	Estar a la altura
Estar a la altura	Meter la pata
	Lanzarse al degüello

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 al 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
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
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
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

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, próximos meses, lacerantes cambios, determinantes medidas, propia alternativa, exitosos frutos, sucesivos meses, estalinista revolución pasadas elecciones, mayor número, siempre oportunista PSN, hasta ahora protegida y saneada Caja de Navarra, antiguas mayorías, mero hecho, buena alumna, mayor burbuja, maldito entramado, legítima mayoría, mordaz pluma

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 incluso han utilizado perífrasis inexistentes en los textos indubitados o 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

CORPUS INDUBITADO	CORPUS SIMULACIÓN
se les debe ayudar –en contadas ocasiones-	debería ser puesto de antemano en conocimiento de la fiscalía
...se les deba ayudar	alguien debe advertir del error
las cosas deben ir	las cosas deben ir por otro lado
la banca debe ser saneada	
debe contar con la disponibilidad	
debiera proscribirse el uso	
debiera gustarles el cambio	
hay que deslegitimar al conjunto del sistema	–
hay que arramplar con la institución	
habrá que imputar el pago	
hemos de mejorar nuestro trabajo cada día	hemos de mejorar nuestro trabajo cada día.
–	el mundo ha de saber de su incompetencia
	tendría que saber
–	deberían de producirse

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 ilustran una preferencia por *-ra-*, y esto ha sido una característica que, aun siendo 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 del subjuntivo

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

7.2.4 Expresión de futuro

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

CUADRO 11. Ejemplos de expresión de futuro

CORPUS INDUBITADO	CORPUS SIMULACIÓN
Habrá que sacar	se irá
Podrá aumentar	tendrá
Habrá que imputar	cooperará
Las cosas podrán cambiar para siempre	te verás
Si será posible	evitaràs
Costará, pero se limpiará	estaran
	harán
	endenderé

Nos van a aportar
Vamos a comprar
Se va a poner orden
Vas a tener

Va a conseguir
Va a creer
va a hacerlo
va a depositar

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 relativas de persona —lo que seguramente es una cuestión idiosincrática, más que fruto de contacto de 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, 17% y 20%), *quien* e incluso ha habido imitadores que han incluido el uso del *cual*.

CUADRO 12. Ejemplos de pronombres relativos

CORPUS INDUBITADO	CORPUS SIMULACIÓN
Quienes actuamos Quienes han tomado Quienes más hablan Para quienes estos días	Quienes más hablan Quienes dijeron
nadie de los que ayer que son menos de los que toman Neptuno	al lado de los que quieren destruir los que sabemos para los que hemos cómplice de los que con su gestión A los que han otorgado
–	Izquierda totalitaria [...], la cual será incapaz me refiero propiamente al cual su partido ninguno de los cuales ha producido momento en el cual pacto [...] al cual pretende añadirse Caja de Navarra La cual podrá ..., lo cual debería

7.3 Pragmático-discursivo

7.3.1 Marcadores discursivos en posición inicial

Los textos indubitados se caracterizan por una escasa presencia de marcadores discursivos y por una escasa variedad de los mismos. En concreto, el autor utiliza para reforzar la argumentación *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 ocasiones no han prestado atención a la posición o a la puntuación. Por 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...).
En efecto, son menos que... sin duda, molestas...	Ellos buscan, sin duda, la ocasión de... sin duda de nuevo...
Buscan, sin duda, la ocasión	sin duda ahora Pues bien, señora mía Es aceptar, sin duda, la ocasión para... Esta es sin duda su actual sin duda pretende

7.3.2 Expresión del énfasis

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

CUADRO 14. Ejemplos de expresión del énfasis mediante negrita

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

7.3.3 Formato de incisos

La expresión de los incisos es otra variable relevante para la descripción del estilo idiolectal en la comparación forense de textos escritos. 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 (—), que es lo 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

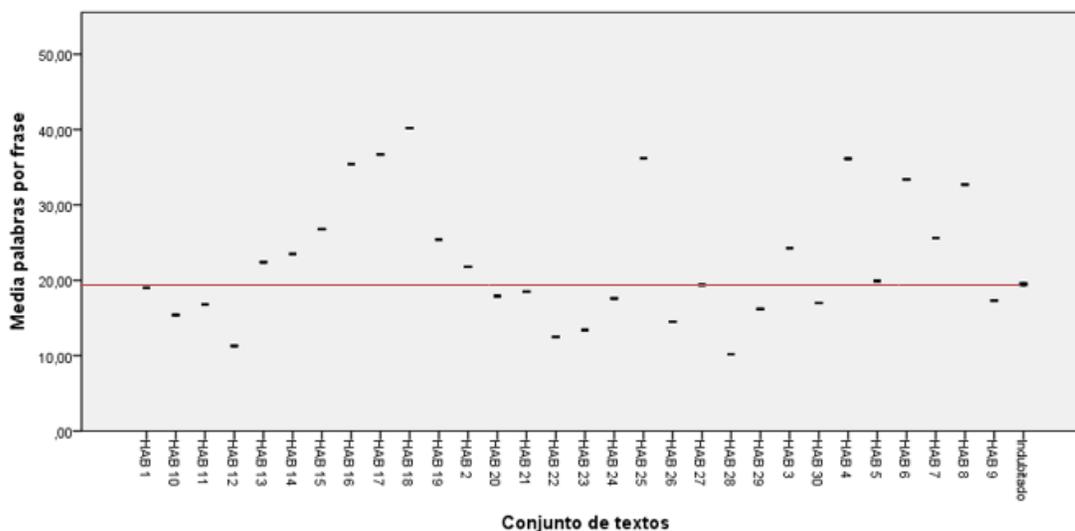
CORPUS INDUBITADO	CORPUS SIMULACIÓN
-es delito transformarlas violentamente- -sin duda, molestas por la reducción de sus libertades- -en contadas ocasiones- (véanse sus soflamas y pasquines) (aunque algunos no quieren enterarse)	– que en este caso la enfermedad parece haber calado hasta lo más hondo de la sociedad- (en efecto, me refiero, propiamente al cual su partido quiere destruir) -Socialista obrero y español recordemos-

7.4 Complejidad

7.4.1 Media de palabras

Dos rasgos que han mostrado ser idiosincráticos son la longitud de frase y de párrafo en términos de palabras por unidad. Ningún imitador ha reproducido con exactitud ni el número medio de palabras por frase, aunque unos pocos se han aproximado. En la Figura 4 se muestra de forma visual los resultados de la media de palabras por frase. Los valores del autor indubitado se muestran en última posición y se ha trazado una línea con su valor para poder visualizar de forma más sencilla qué autores son más capaces de reproducir el rasgo.

FIGURA 4. Media de palabras por frase



Se procede a realizar la prueba de contraste t con un valor de prueba 19.35 correspondiente al valor medio de los textos indubitados 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

Valor de prueba = 19.35

	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Media de palabras por frase	2.238	29	.033	3.47667	.2989	6.6544

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 un 86.7%, 83.3% y 66.7% de éxito en la imitación, respectivamente. Por el contrario, los rasgos imposibles de emular para los imitadores de este estudio han sido principalmente las variables morfosintácticas (los pronombres relativos, la perífrasis de obligación, el futuro) y léxicas (los anglicismos y las expresiones). Con un bajo porcentaje de éxito encontramos las variables de complejidad (la media de palabras por frase), así como las expresiones, los marcadores discursivos, el adjetivo antepuesto y el formato de incisos. En el Cuadro 16 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
Frases 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 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 únicamente presentes en algunos de los textos del corpus de simulación. La clase de errores normativos que comete un individuo puede dar pistas muy importantes sobre cuestiones sociolingüísticas relacionadas con su nivel educativo, el contacto de lenguas o la familiaridad del individuo en la redacción de ese tipo de textos. Algunas de los errores detectados en el corpus de simulación y que nos indicaría 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 incisos) 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 Sr. Cervera, 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 metódico.

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 idiosincrático del corpus indubitado, lo cual indica que, efectivamente, el proceso de imitación requiere entre otros de un gran dominio lingüístico y capacidad de atención a la forma y 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 de su propio estilo idiolectal las variables que detectaron en el corpus indubitado (p. ej. 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 marcadez 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 de Santiago Cervera 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 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 irrepetible, no se compone únicamente de una variable, sino que es precisamente todo el conjunto de variables analizadas lo que caracterizan el uso del lenguaje particular y único de un autor determinado. De este modo, se ha podido comprobar que la naturaleza de un discurso está configurada por toda una serie de rasgos cuya concurrencia en un texto pasa inadvertida a cualquier usuario de la lengua hablada y no se puede imitar con un nivel de precisión suficiente para no ser detectado en un análisis de comparación de textos exhaustivo, 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. Los imitadores de este estudio 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. Se ha podido comprobar que el estilo idiolectal de un individuo es tan inherente que marca cada discurso que realiza.

En conclusión, los resultados de este estudio permiten refutar el recurrente argumento a favor de la posibilidad de imitación del estilo del autor. Se ha demostrado que la suplantación del estilo del autor es difícil de realizar y, además, que 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*, XXVII, 2, 1–18.

Asociación de Academias de la Lengua Española (ASALE). (2005). *Diccionario panhispánico de dudas*. Recuperado de <http://lema.rae.es/dpd>. Consultado 16 mayo de 2017.

BUTLER, CHRIS. (1998). Enriching the FG lexicon. En Hella Olbertz, Kees Hangeveld y Jesús Sánchez (Eds.), *The Structure of the Lexicon in Functional Grammar* (pp.171–194). Amsterdam/Philadelphia: John Benjamins,

CANTER, DAVID, & CHESTER, JOANNE. (1997). Investigation into the claim of weighted Cusum in authorship attribution studies. *Forensic Linguistics*, 4, 252–261.

CERVERA SOTO, SANTIAGO. (2012). Blog de Santiago Cervera. Recuperado de <http://santiagocervera.blogspot.com.es/>

CHASKI, CAROLE. (2001). Empirical Evaluations of Language-Based Author Identification Techniques. *International Journal of Speech Language and the Law*, 8(1), 1–57.

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*. London/New 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.

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? In Coulthard, Malcolm & Johnson, Alison. (Eds.). (pp. 508–522). *The Routledge Handbook of Forensic Linguistics*. Abingdom: Routledge.

HERNÁNDEZ ESTEBAN, MIREIA. (2016). *Lingüística Forense Básica*. Tirant lo Blanch.

HOWALD, BLAKE STEPHEN. (2009). Authorship attribution under the rules of evidence: empirical approaches in the layperson legal system. *The International Journal of Forensic Linguistics*, 15(2), 219–247.

JAKOBSON, ROMAN. (1971 [1956]). Two aspects of language and two types of aphasic disturbances. En *Selected Writings II: Word and Language*. The Hague: Mouton.

JOHNSTONE, BARBARA. (2000). The Individual Voice in Language. *Annual Review of Anthropology*, 29, 405–424.

MCMENAMIN, GERALD. (2001). Style markers in authorship studies. *Forensic Linguistics* 8(2), 93–97.

QUERALT, SHEILA. (2014). Acerca de la prueba lingüística en atribución de autoría hoy. *Revista de Llengua i Dret*, 62, 34–48.

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*. Pearson educación.

TURELL, MARIA TERESA. (2010). The use of textual, gramatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech, Language and the Law*, 17, 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, 67–83.

WRIGHT, DAVID. (2014) *Stylistics versus Statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails*. (Tesis doctoral, University of Leeds). Recuperada de <http://etheses.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 cybercrime investigation. *ISI'03 Proceedings of the 1st NSF/NIJ Conference on Intelligence and Security Informatics*, 59–73.