

ESTABILIDAD IDIOLECTAL DEL ESPAÑOL A TRAVÉS DE CUATRO GÉNEROS DE COMUNICACIÓN: APORTACIONES AL ANÁLISIS DE AUTORÍA FORENSE

Andrea Mojedano Batel, Neus Alberich Buera, Krzysztof Kredens*

Resumen

El análisis de autoría forense se basa en que cada hablante tiene su propia versión de una lengua o *idiolecto* cuyos rasgos lingüísticos son recurrentes de manera relativamente estable (Coulthard et al., 2011). En lingüística, hay una gran disparidad entre el conocimiento del concepto del idiolecto y su estudio con datos empíricos (Barlow, 2013). En esta investigación multigenérica (*cross-genre*, en inglés) cubrimos parte de estas lagunas, con datos de corpus escritos y orales del español mexicano. A través de un análisis de n-gramas de palabra, identificamos cuatro áreas de estabilidad idiolectal en español: lenguaje evaluativo y expresiones de cantidad, de modalidad deóntica y de modalidad epistémica. Nuestros resultados muestran que las construcciones gramaticales son útiles en el análisis de autoría, considerando el contexto forense, con análisis motivados lingüísticamente combinados con herramientas cuantitativas. Asimismo, proporcionamos un análisis lingüístico fundamentado en la teoría de la gramática de construcciones basada en el uso, la cual da cuenta adecuadamente del uso recurrente y particular de las construcciones idiolectales.

Palabras clave: idiolecto; análisis multigenérico; variación idiolectal; análisis de autoría; lingüística forense.

IDIOLECTAL STABILITY OF SPANISH ACROSS FOUR COMMUNICATIVE GENRES: CONTRIBUTIONS TO FORENSIC AUTHORSHIP ANALYSIS

Abstract

Forensic authorship analysis is based on the assumption that each speaker has his or her own version of a language, or idiolect, whose linguistic features show a relatively stable recurrence (Coulthard et al., 2011). In linguistics, there is a great disparity between knowledge of the concept of idiolect and its study using empirical data (Barlow, 2013). In this cross-genre study, we fill in some of these gaps with data from written and spoken corpora of Mexican Spanish. By analysing word n-grams, we identify four areas of idiolectal stability in Spanish: evaluative language and expressions of quantity, deontic modality and epistemic modality. Our results show that grammatical constructions are useful in authorship analysis, considering the forensic context, combining linguistically reasoned analyses with quantitative tools. We also provide a linguistic analysis grounded on the usage-based construction grammar theory, which adequately accounts for the recurrent personal use of idiolectal constructions.

Keywords: idiolect; cross-genre analysis; idiolectal variation; authorship analysis; forensic linguistics.

* Andrea Mojedano Batel, investigadora posdoctoral como parte del proyecto de *Expanding Excellence in England* (E3) de *Research England* en Lingüística Forense en el Instituto Aston para la Lingüística Forense (AIFL) de la Universidad de Aston (Inglaterra) y miembro del Centro para el Análisis de Textos Forenses del AIFL. a.mojedanobatel@aston.ac.uk. [id 0000-0001-9028-8279](https://orcid.org/0000-0001-9028-8279).

Neus Alberich Buera, estudiante de doctorado en Lingüística Forense en la Universidad de Aston (Inglaterra). 190184416@aston.ac.uk.

Krzysztof Kredens, director del programa de máster en Lingüística Forense y de los programas de posgrado de inglés de la Universidad de Aston (Inglaterra) y director del Centro para el Análisis de Textos Forenses del Instituto Aston para la Lingüística Forense. kredenkj@aston.ac.uk. [id 0000-0001-7038-9478](https://orcid.org/0000-0001-7038-9478).

Artículo recibido el 31.01.2023. Evaluaciones ciegas: 09.02.2023 y 23.02.2023. Aceptación de la versión final: 14.03.2022.

Citación recomendada: Mojedano Batel, Andrea, Alberich Buera, Neus, y Kredens, Krzysztof. (2023). Estabilidad idiolectal del español a través de cuatro géneros de comunicación: aportaciones al análisis de autoría forense. *Revista de Llengua i Dret, Journal of Language and Law*, 79, 285-304. <https://doi.org/10.58992/rld.i79.2023.3951>

Sumario

1 Introducción y estudios previos

2 Marco teórico

3 Participantes

4 El corpus

5 Marco metodológico

5.1 Métodos cuantitativos de análisis

5.1.1 Extracción computacional de construcciones gramaticales adecuadas

5.1.2 Determinación del nivel de regularidad y de peculiaridad de las construcciones lingüísticas

5.2 Método cualitativo de análisis

6 Resultados y discusión

6.1 Lenguaje evaluativo

6.2 Expresión de la modalidad

6.2.1 Expresión de la modalidad deóntica

6.2.2 Expresión de la modalidad epistémica

6.3 Expresiones de cantidad

7 Conclusiones y direcciones futuras

8 Referencias

1 Introducción y estudios previos

El presente estudio examina cuatro áreas lingüísticas que muestran estabilidad idiolectal a través de distintos géneros de comunicación, con datos provenientes de nueve participantes mexicanos pertenecientes a una misma comunidad de habla. El estudio contribuye tanto al desarrollo de una teoría empíricamente validada sobre la noción de idiolecto como a los estudios de análisis de autoría forense con datos en español, pues, aunque el término *idiolecto* es bien conocido en lingüística, hay gran disparidad entre el conocimiento del concepto y el estudio del fenómeno con datos empíricos (Barlow, 2013). Los resultados de esta investigación pueden aplicarse a una amplia gama de tareas forenses, ya que, al identificar y describir áreas lingüísticas estables con sus correspondientes construcciones gramaticales a través de distintos géneros de comunicación, será razonable argumentar que esas mismas áreas lingüísticas y construcciones gramaticales también aparecerán cuando se traten géneros más similares entre sí. Este estudio constituye uno de los primeros en el ámbito multigenérico (en inglés, *cross-genre*) que analiza datos lingüísticos del español.

La noción de análisis de autoría se basa en dos premisas: por un lado, que cada usuario tiene su propia versión de una lengua, o idiolecto (Bloch, 1948), y, por otro, que los rasgos lingüísticos que caracterizan el uso individual de la lengua son recurrentes de manera más o menos estable en el habla, de modo que son las construcciones lingüísticas similares las que caracterizan las distintas producciones lingüísticas de un mismo autor (Coulthard et al., 2011; Grant, 2022).¹ No obstante, algunas investigaciones sugieren que los usuarios tienden a acomodar el habla según el contexto situacional en el que se esté dando el acto comunicativo y el público al que vaya dirigido (Giles y Powesland, 1975; Bell, 1984; Rickford y McNair-Knox, 1994). Esta variación lingüística en un mismo usuario, también llamada *variación intrausuario*, pone en entredicho la estabilidad idiolectal a través de distintos géneros comunicativos.

Numerosos estudios han propuesto distintos métodos y rasgos lingüísticos apropiados para el análisis de autoría (p. ej., Stamatatos, 2013 y 2017; Wright, 2014 y 2017). Sin embargo, estos trabajos solo examinan materiales sociolingüísticos homogéneos. Este estudio, al analizar datos sociolingüísticos heterogéneos, pretende hacer avanzar el campo del análisis de autoría forense.

Dentro de los escasos estudios multigenéricos, Goldstein-Stewart et al. (2009) investigan la diferenciación de usuarios de una lengua en distintos géneros de comunicación, aunque no dan cuenta de los rasgos lingüísticos idiolectales. Este también es el caso de Stamatatos (2013), que emplea un método basado en el número de caracteres, una propiedad textual que no puede ser explicada desde la lingüística. Otros estudios analizan datos que son demasiado parecidos entre sí y que no constituyen distintos géneros de comunicación. La mayoría de las investigaciones se centran en producciones escritas, de manera que no se incluyen distintos modos de producción lingüística (p. ej., Baayen et al., 2002; Spassova, 2009; Overdorf y Greenstadt, 2016; Johnson y Wright, 2017; Litvinova et al., 2018), o emplean datos escritos y orales pero que no difieren en su propósito (p. ej., Litvinova, Litvinova y Seredin, 2018). Asimismo, con la excepción de los estudios de Litvinova y colegas, quienes analizan datos en ruso, casi todos los estudios multigenéricos de variación idiolectal se han llevado a cabo con datos en inglés. La única excepción, a nuestro saber, es el trabajo sobre el español de Mojedano Batel et al. (en prensa), cuyos resultados muestran que hay estabilidad idiolectal multigenérica en las construcciones intensificadoras (p. ej., “qué” + adjetivo), cuantificadoras (p. ej., “un poco”) y de modalidad deóntica (p. ej., “tener que” + cláusula) y epistémica (p. ej., “veo que” + cláusula, o partículas discursivas como “o sea”). Mojedano Batel et al. (en prensa) se enfoca únicamente en el análisis de las construcciones de modalidad epistémica. El presente estudio, que parte de aquel, pretende analizar todas las áreas idiolectales estables de manera global, aportando nuevos análisis en el área de modalidad epistémica.

En el marco de la sociolingüística hispánica, los estudios de variación idiolectal han analizado elementos lingüísticos concretos, como los marcadores del discurso en narraciones (Romano y Porto, 2013), el uso de las formas adverbiales “quizá(s)” y “tal vez” (Troya, 2015), la posición de clíticos (Iglesias, 2016) o el uso de “que” en cláusulas subordinadas por “creer”, el uso de la preposición “de” en perífrasis modales con “deber” y la alternancia entre “donde” y “que/cual” como nexos (Blas, 2017). Los resultados de estos trabajos indican

¹ Las *construcciones*, siguiendo a Goldberg (2003, p. 219), son pareados de forma y función almacenados en la mente del usuario, que incluyen morfemas, palabras y patrones lingüísticos.

que algunos rasgos lingüísticos se pueden relacionar con las preferencias idiosincráticas de los usuarios de una lengua.

La meta principal de nuestro estudio es investigar y describir la estabilidad idiolectal en español con datos lingüísticos heterogéneos provenientes de cuatro géneros de comunicación. Nuestro trabajo se ancla en el estudio del lenguaje probatorio o evidencial, una de las áreas principales de lingüística forense (Gibbons y Turell, 2008), centrándose en el subcampo del análisis de autoría, con la intención de avanzar el conocimiento lingüístico en español para asesorar en investigaciones y contribuir a la obtención de pruebas en casos judiciales. La lingüística forense se ha venido desarrollando en España en los últimos años, donde tiene ya una sólida presencia social y donde algunos investigadores en este campo han trabajado el tema de análisis de autoría (p. ej., Turell, 2011; Queralt, 2014; Queralt y Giménez, 2018; Cicres y Queralt, 2019) y de variación idiolectal, donde tanto Turell y Gavalda (2013) como Spassova et al. (2014) proponen un índice de similitud o distancia idiolectal, que pretende calcular la similitud entre producciones lingüísticas para determinar si proceden de un mismo individuo, basándose en variables lingüísticas discretas. Dicho esto, el tema de estabilidad idiolectal a través de distintos géneros es uno que apenas se investiga, aunque los hallazgos sobre el mismo son vitales para las tareas forenses, y nuestro estudio pretende llenar parte de las lagunas sobre el tema poniendo el foco en el español, donde, a nuestro saber, solo hay un estudio previo, el de Mojedano Batel et al. (en prensa).

En el presente trabajo entendemos *un género de comunicación* de la manera tradicional, en términos de un propósito compartido y convenciones comunes de contenido y estilo (Swales, 1990). Al haber identificado la existencia de estabilidad idiolectal multigenérica en español, y, más específicamente, en construcciones intensificadoras, de cuantificación, y de modalidad deóntica y epistémica (Mojedano Batel et al., en prensa), este estudio ofrece una descripción general de todos estos tipos de construcciones con datos naturales multigenéricos del español mexicano, proporcionando información sobre las estrategias de estabilidad idiolectal de varios individuos en una misma comunidad de habla. Esta descripción toma en cuenta el contexto de los peritos judiciales en lingüística forense, quienes prefieren métodos que emplean análisis motivados lingüísticamente, combinados con herramientas cuantitativas (Solan y Tiersma, 2004, p. 463).

Partimos de la premisa de que el análisis de autoría debe poder explicarse desde una teoría de la variación lingüística, dado que un solo individuo –generalmente situado en la periferia de sus redes sociales– puede ser el responsable de procesos clave en el cambio lingüístico (Milroy, 1992). Así, nuestro trabajo se enmarca tanto en el campo de la lingüística forense como en el de la sociolingüística. La lingüística forense sitúa al individuo en un primer plano, alineándose con la sociolingüística variacionista de la tercera ola (*third wave variationist sociolinguistics*, en inglés). En esta rama de la sociolingüística, la lengua y la identidad son inseparables: la lengua no es un reflejo de la identidad, sino que se utiliza para constituirla. Esta identidad puede surgir de varias maneras, que van desde la acción deliberada hasta la práctica habitual, y también como resultado de procesos de interacción entre usuarios de una lengua (Bucholtz y Hall, 2010). Siguiendo a Drummond y Schleff (2016), esto indica que un cambio en la identidad de alguien puede resultar en un cambio en su práctica lingüística, y entonces para estudiar y comprender la identidad debemos no solo centrarnos en una característica lingüística particular, sino en algo mayor, que, para muchos variacionistas, implica examinar tanto el estilo, es decir, los rasgos fonológicos y morfosintácticos (Eckert, 2000; Moore, 2011), como las orientaciones evaluativas, afectivas y epistémicas del discurso (Ochs, 1992; Rauniomaa, 2003; Kiesling, 2013). Aquí revisamos rasgos pertenecientes a estas dos macroáreas para analizar la estabilidad idiolectal.

2 Marco teórico

Nos acercamos al estudio de la estabilidad idiolectal a través de la gramática de construcciones basadas en el uso. Esta teoría (*usage-based construction grammar*, en inglés) concibe la gramática como un producto de la experiencia (Hopper, 1987). Entendemos las *construcciones* como pareados de forma y función almacenados en la mente del usuario de una lengua (de aquí en adelante, “usuario”), que incluyen morfemas, palabras y patrones lingüísticos tanto léxicamente semirrellenos como completamente generales (Goldberg, 2003, p. 219). Siguiendo a Bybee y Eddington (2006), también consideramos construcciones a estos pareados incluso cuando su forma, función y/o significado es totalmente predecible. Este enfoque nos permite comprender cómo surgen construcciones con características impredecibles, como las funciones pragmáticas especiales. Dentro de la

gramática de construcciones basadas en el uso, la noción de *arraigo* (*entrenchment*, en inglés), desarrollada por Langacker (1987), se refiere al nivel en el que la formación y activación de una unidad cognitiva se automatiza a través de la rutina y la repetición en la mente del hablante, hasta el punto de convertirse en una unidad; por otra parte, las unidades se arraigan más o menos según la frecuencia de uso de las mismas.

En el contexto forense, Wright (2017) anota que hay secuencias de palabras (o construcciones lingüísticas) específicas que están funcionalmente unidas a contextos, rutinas y propósitos comunicativos concretos y recurrentes. Mientras que algunas de estas rutinas y sus construcciones lingüísticas resultantes son parte de la producción lingüística de la comunidad de habla en general, algunas otras construcciones son más personales, generadas solo por un miembro de la comunidad de habla en cuestión, aspecto que ya antes había notado Mufwene (2010) para la lingüística general. Wright mantiene que las construcciones ofrecen al analista de autoría un rasgo lingüístico –para el que hay consenso teórico basado en la noción de arraigo– que puede ayudar a explicar diferencias idiolectales entre autores.

En el presente estudio partimos de la premisa de que la frecuencia multigenérica de uso de una construcción, es decir, el número de géneros de comunicación en el que una construcción aparece, también puede indicar el arraigo de dicha estructura en la mente del usuario de la lengua. Entendemos que mientras más géneros haya donde aparece una construcción específica para un usuario en concreto, más arraigada estará esta construcción en la mente del mismo, ya que, como se ha mencionado antes, los usuarios tienden a modificar el habla según el contexto situacional en el que se encuentren y el público al que se dirijan (Giles y Powesland, 1975; Bell, 1984; Rickford y McNair-Knox, 1994): deducimos que la estabilidad en el uso de construcciones a pesar de diferencias de género y modo muestra un arraigo mental de las mismas. Nuestra investigación se enfoca en frecuencias multigenéricas (es decir, a través de diferentes géneros), dejando para estudios futuros otro tipo de enfoques, como el análisis de la frecuencia de uso de una construcción en términos del número total de instancias.

3 Participantes

Seleccionamos a los participantes del estudio basándonos en la homogeneidad de su trasfondo social; todos provienen de una misma comunidad de habla. Estos participantes se dividen en seis mujeres y tres hombres de entre 30 y 60 años en el momento de ser entrevistados (en el año 2020), y provienen del área central de México. Todos son profesores o investigadores en una universidad pública de la Ciudad de México, donde viven, y su lengua materna es el español. Ninguno de los participantes es bilingüe, aunque en este estudio no recopilamos más información que esta sobre conocimiento de segundas lenguas.²

El hecho de que los participantes del estudio formen parte de una misma comunidad de habla es vital, ya que, como anota Grant (2010), es difícil distinguir la producción lingüística individual de la grupal en una comunidad de habla: al examinar idiolectos provenientes de una misma comunidad de habla, pretendemos ver cuáles son los rasgos lingüísticos que pueden, potencialmente, identificar a usuarios con un trasfondo social similar. Para llevar a cabo este acercamiento al tema de la estabilidad idiolectal, siendo este uno de los primeros estudios que analizan el español, examinaremos aquellas construcciones de los nueve usuarios que hayan aparecido en los cuatro géneros del corpus.

4 El corpus

De cada uno de los participantes obtuvimos los siguientes datos lingüísticos, que anonimizamos debidamente:³

1. 30 mensajes de texto (en concreto, mensajes de WhatsApp), con un total de 6122 palabras. Todos los mensajes de texto datan de 2020.
2. 30 correos electrónicos, con un total de 33203 palabras. Todos datan de 2019 o 2020.

2 Entendemos la adquisición bilingüe como una exposición regular a dos o más lenguas desde el nacimiento hasta antes de los dos años, siguiendo a De Houwer (2005).

3 Este mismo corpus es el que se usó en un estudio inicial, el de Mojedano Batel et al. (en prensa), que dio lugar a este segundo análisis.

3. Una entrevista sociolingüística semidirigida con cada uno de los participantes, realizada por una de las investigadoras del presente estudio entre 2019 y 2020, con un total de 18842 palabras y 174 minutos de audio. La temática principal de la conversación se basó en preguntar a los participantes sobre su comida favorita y, después, dejar que hablaran sobre cualquier tema que les interesara, haciendo preguntas para continuar con la conversación, en línea con el principio de desplazamiento tangencial (Labov, 1984) como una manera de reducir al máximo la paradoja del observador (Labov, 1972).
4. Transcripciones de varias reuniones mensuales de trabajo donde participaron los voluntarios, con un total de 844104 palabras. Estas transcripciones se encuentran en el dominio público y son totalmente accesibles a través de la web. Las reuniones tuvieron lugar entre los años 2008 y 2019. En dichas juntas, se tratan temas estrictamente profesionales, relacionados con la universidad y su funcionamiento.

La tabla 1 muestra el número total de palabras por cada usuario y género comunicativo en nuestro corpus.

Tabla 1

Número de palabras por usuario y género comunicativo

	Mensajes de texto: número de palabras	Correos electrónicos: número de palabras	Entrevista sociolingüística: número de palabras	Reuniones de trabajo: número de palabras	Número total de palabras
Usuaría 1: mujer	326	1179	2919	65858	70279
Usuaría 2: mujer	503	6549	2148	81560	90760
Usuario 3: hombre	270	3185	4259	34297	42011
Usuaría 4: mujer	2114	5306	1562	19980	28962
Usuaría 5: mujer	856	4916	956	1266	7994⁴
Usuario 6: hombre	397	3114	1104	429681	434296
Usuario 7: hombre	332	1562	1387	146647	149928
Usuaría 8: mujer	636	6388	3110	47056	57190
Usuaría 9: mujer	688	1004	1397	17759	20848
Total	6122	33203	18842	844104	902271

Fuente. Elaboración propia.

Los datos se produjeron en contextos naturales y proceden de distintos géneros, lo que nos permite comprobar la estabilidad de estilos idiolectales en situaciones comunicativas multimodales. La tabla 2 ofrece un resumen de los datos del corpus con sus respectivas características comunicativas y variación diafásica relacionada con la adscripción al registro de cada género.

Tabla 2

Resumen de los datos con sus respectivas características comunicativas

Género	Modo de producción	Formalidad	Público
Mensajes de texto	Escrito (teclado del teléfono)	Informal	Socialmente cercano
Correos electrónicos	Escrito (teclado del ordenador)	Formal	Socialmente semidistante
Entrevista	Hablado (interacción entre dos usuarios)	Informal	Socialmente semidistante
Reuniones de trabajo	Hablado (interacción grupal)	Formal	Socialmente distante

Fuente. Elaboración propia.

5 Marco metodológico

Atendiendo las demandas metodológicas de las pericias lingüísticas en el ámbito judicial, este trabajo sigue una metodología mixta, con métodos cuantitativos para dar objetividad a las construcciones gramaticales que se analicen y métodos cualitativos para dar fundamentación lingüística a los resultados que indiquen los modelos computacionales cuantitativos. En el campo de la lingüística forense, esta metodología está respaldada por Solan y Tiersma (2004) y ha sido aplicada por Johnson y Wright (2017).

Los métodos cuantitativos (sección 5.1) incluyen la extracción computacional de palabras y la determinación de los niveles de regularidad y peculiaridad de las construcciones de cada usuario: esta metodología cuantitativa ya ha sido descrita previamente en Mojedano Batel et al. (en prensa), aunque difiere de aquel estudio inicial al solo analizar construcciones que aparecen en todos los géneros posibles. El método cualitativo (sección 5.2) es completamente nuevo y, por ende, diferente del de nuestro estudio inicial, enfocándose en analizar todas las construcciones que muestran estabilidad en los cuatro géneros comunicativos estudiados, aunque poniendo especial atención en las construcciones con mayor potencial diferenciador entre idiolectos.

5.1 Métodos cuantitativos de análisis

5.1.1 Extracción computacional de construcciones gramaticales adecuadas

Las construcciones que analizamos en el estudio son *n*-gramas de palabras. Johnson y Wright (2017) indican que las secuencias de palabras (que son sinónimas a las construcciones gramaticales) constituyen rasgos lingüísticos fácilmente identificables y accesibles que dan cuenta de preferencias léxicas y sintácticas individuales. Asimismo, dado que hay explicaciones teóricas de las razones por las cuales los usuarios varían en el uso de construcciones lingüísticas (véase la sección 2), las construcciones gramaticales son ideales para el análisis de autoría forense. Un *n*-grama de palabras es un término operativo que se refiere a cualquier secuencia de *n* número de palabras de largo (Juola, 2008), y en el presente estudio analizamos *n*-gramas de una a cuatro palabras de largo. Es importante especificar que fundamentamos el análisis en formas lingüísticas lematizadas para poder tener una visión más amplia de los posibles patrones de uso lingüístico individuales. Cuando fue necesario, se examinaron las formas declinadas o conjugadas de los lemas.

Mediante herramientas computacionales obtuvimos una lista de todos los *n*-gramas que ocurrían en los cuatro géneros de producción lingüística de cada participante. Posteriormente, prescindimos: (a) de todas las palabras funcionales altamente frecuentes en lengua española, tales como preposiciones o artículos, por su bajo potencial discriminatorio, y (b) de las construcciones que aparecían a causa del tema que se discutía y que, por tanto, no arrojaban luz sobre los posibles hábitos lingüísticos interiorizados de los usuarios. Es cierto que, cuando las palabras funcionales formaban parte de construcciones más largas, estas sí se tenían en cuenta. Por ejemplo, descartamos ocurrencias de la preposición “a”, pero examinamos construcciones como “a la mejor”.

5.1.2 Determinación del nivel de regularidad y de peculiaridad de las construcciones lingüísticas

Grant (2010) establece que para la resolución de casos forenses es necesaria una metodología que permita determinar tanto el nivel de regularidad de las características lingüísticas entre las distintas producciones de un usuario (*consistency*, en inglés) como el nivel de peculiaridad de esas características lingüísticas entre usuarios (*distinctiveness*, en inglés). Esto es, si hay construcciones que parecen estables en el habla de un individuo, debemos examinar si estas son lo suficientemente peculiares como para poder distinguir entre diferentes autores.

Para hacer de esta una metodología operativa, usamos un recuento de géneros (de aquí en adelante, RG) como medida básica para dar cuenta del número de géneros en los que un participante produce una construcción lingüística. Como ejemplo, la usuaria 1 produjo el bigrama “creer que” 62 veces en dos géneros. Entonces, la frecuencia total de “creer que” para esta usuaria es de 62, y el valor del RG de este bigrama, en el que enfocamos nuestro análisis, es de 2. A modo de comparación, la usuaria 2 produjo el bigrama “creer que” 191 veces en todos los géneros posibles, es decir, en cuatro géneros. La frecuencia total de “creer que” para la usuaria 2 es de 191, y el valor del RG de este bigrama es 4. El recuento de géneros nos permite evaluar la estabilidad de las construcciones de un usuario de manera multigenérica, es decir, a través de los cuatro géneros de comunicación. El RG es la medida básica de frecuencia en la que nos centramos en el presente estudio. La tabla 3 ofrece esta información sobre “creer que” de manera organizada, de modo que permite al lector apreciarla más rápidamente.

Tabla 3

Frecuencias y distribución de “creer que” en los subcorpus de las usuarias 1 y 2

Tipo de n-grama	usuario	Construcción	Géneros comunicativos	Recuento de géneros (RG)	Frecuencia total
bigrama	usuaria 1	creer que	entrevista 1	2	62
			reuniones 61		
bigrama	usuaria 2	creer que	mensaje de texto 2	4	191
			reuniones 176		
			e-mails 10		
			entrevista 3		

Fuente. Elaboración propia.

En esta investigación nos enfocamos en el análisis de construcciones con un RG de 4, es decir, en construcciones que aparezcan en todos los géneros posibles en el habla de cada participante. Sin embargo, es importante hacer el recuento de géneros, puesto que, si una construcción con un RG de 4 aparece en más de un usuario, tendremos que ver si esta construcción es potencialmente diferenciadora de un idiolecto o si, quizás, muestra más bien una tendencia a ser una construcción común, ampliamente adoptada por varios miembros de una misma comunidad de habla. Para poder hacer operativo este concepto, comparamos el RG de un usuario con la media de la RG calculada para los ocho usuarios restantes. Como ejemplo, el RG de “creer que” para la usuaria 2 es de 4, mientras que la media del recuento de géneros calculada para el resto de los usuarios (de aquí en adelante, MRG) es de 3,25. Esto indica que la usuaria 2 produjo el bigrama “creer que” en los cuatro géneros, mientras que los ocho participantes restantes produjeron “creer que” en una media de más de tres géneros. Esto nos dice que el bigrama “creer que” es común en la mayoría de los idiolectos del corpus y, por tanto, no es potencialmente diferenciador de un idiolecto. La tabla 4 provee una lectura rápida y eficiente de la comparación de la MRG de “creer que”, que previamente discutimos.

Tabla 4

Comparación de la MRG de “creer que” entre la usuaria 2 y el resto de los usuarios⁴

Tipo de n-grama	Usuario	Construcción	Géneros comunicativos	Recuento de géneros (RG)	Frecuencia total
bigrama	usuaria 2	creer que	U2 = 4	4	191
bigrama	usuarios 1, 3, 4, 5, 6, 7, 8, 9	creer que	U1 = 2, U3 = 3, U4 = 4, U5 = 3, U6 = 4, U7 = 4, U8 = 3, U9 = 3	(MRG) 3,25	1377

Fuente. Elaboración propia.

En resumen, una MRG relativamente alta (más cerca de cuatro que de cero) indicará que la construcción en cuestión posiblemente no sea idiolectal, mientras que una MRG relativamente baja (más cerca de cero que de cuatro) apuntará a una construcción probablemente idiolectal. Esta aproximación permite evaluar la estabilidad de las construcciones gramaticales en las distintas producciones de un mismo usuario y posibilita la consideración de los patrones de uso de un usuario en relación con el resto de los usuarios.

5.2 Método cualitativo de análisis

Una vez delimitadas las construcciones relevantes, examinamos cualitativamente su uso contextual, obteniendo las construcciones y su contexto mediante el programa AntConc (Anthony, 2017), con el fin de encontrar patrones de uso lingüístico estables en el idiolecto de cada participante. Examinamos todas aquellas construcciones que muestran un RG de 4, sin importar su MRG, es decir, sin considerar si muestran potencial o no a la hora de diferenciar idiolectos. De esta manera, podremos vislumbrar el comportamiento general de una comunidad de habla, observando: (a) qué construcciones específicas tienden a ser compartidas por toda o gran parte de la comunidad y qué construcciones tienden a ser producidas por pocos miembros de la misma, y (b) si hay elementos sociolingüísticos comunes en aquellos miembros de la comunidad de habla que muestran un uso idiosincrático del lenguaje a través de distintos géneros. Dicho esto, pondremos especial énfasis en las construcciones que muestran una MRG menor a 2,5, es decir, en aquellas construcciones con mayor potencial diferenciador entre idiolectos.

6 Resultados y discusión

El análisis cuantitativo permitió vislumbrar las construcciones que ocurren por lo menos una vez en cada uno de los géneros analizados por usuario, es decir, las construcciones que muestran un RG de 4. Estas construcciones pertenecen a cuatro áreas de funciones lingüísticas: (1) lenguaje evaluativo, (2) marcadores de modalidad deóntica, (3) marcadores de modalidad epistémica y (4) expresiones de cantidad. A continuación, discutimos estas construcciones dividiéndolas por áreas de función lingüística, evaluando su potencial discriminatorio en el contexto del análisis de autoría forense.

6.1 Lenguaje evaluativo

La evaluación en un discurso es un concepto escurridizo, difícil de describir. Sin embargo, en el presente trabajo seguimos la definición de Thompson y Hunston (2000, p. 5), quienes describen la *evaluación* como la expresión de la actitud o punto de vista del usuario sobre las entidades o proposiciones sobre las que está hablando. Esta definición puede, de manera problemática, superponerse parcialmente a la definición de *epistemicidad* que ofrecemos en la sección 6.2; sin embargo, aquí delimitamos el *lenguaje evaluativo* a aquellas valoraciones donde los usuarios perciben una entidad o proposición como positiva o negativa, obviando cualquier otro tipo de evaluación.

Dentro del lenguaje evaluativo, nuestros datos muestran que las construcciones [“muy” + adjetivo calificativo] (1) y [“muy” + adverbio de modo] (2) constituyen un rasgo destacado (RG = 4) en la producción lingüística

⁴ Abreviamos los términos “usuaria 1”, “usuaria 2”, etcétera, como “U1”, “U2”, y así sucesivamente.

de cuatro usuarios del corpus (U1, U2, U6 y U8). Estos usuarios usan “muy” para intensificar tanto adjetivos como adverbios, es decir, en su habla es típica no solo la evaluación, sino una evaluación intensificada.

- | | | | |
|-----|----|---|-------------------------------|
| (1) | a. | esto [...] se ha vuelto muy complicado | (usuaria 2; entrevista) |
| | b. | es un clima [...] muy bueno | (usuario 6; entrevista) |
| | c. | se trata de una cuestión muy seria | (usuaria 8; reunión) |
| (2) | a. | me parece [...] muy bien | (usuaria 1; mensaje de texto) |

Hemos de notar que el uso de estos dos tipos de construcciones no ofrece potencial discriminatorio, ya que todos los usuarios del corpus acuden al mismo recurso para indicar una intensificación de la cualidad descrita por el adjetivo o adverbio que sigue a “muy” (MRG = 3). La diferencia entre los usuarios 1, 2, 6 y 8 y los cinco usuarios restantes es que los primeros usan estas construcciones en todos los géneros estudiados, mientras que los segundos lo hacen solo en algunos de los géneros.

Sin embargo, en nuestro análisis se observa el uso de otra construcción que permite expresar una evaluación tanto positiva como negativa: la construcción formada por el adjetivo exclamativo “qué” seguido por un adjetivo que permite evaluar o enfatizar ciertas entidades o proposiciones (3-4). La construcción [“qué” + adjetivo] puede, en ocasiones (3), funcionar como un adverbio de afirmación. Esta construcción solo fue producida por la usuaria 8 con un RG de 4, hecho que sugiere que esta usuaria ha desarrollado un hábito lingüístico poco común, dada la media inferior a dos en el recuento de géneros (MRG = 1,63) de esta construcción.

- | | | | |
|-----|----|------------------------------------|-------------------------------|
| (3) | a. | qué bueno que habló contigo | (usuaria 8; mensaje de texto) |
| | d. | qué bueno que me lo aclaras | (usuaria 8; reunión) |
| (4) | b. | qué mala noticia | (usuaria 8; e-mail) |
| | c. | ay, ¡qué lindo! | (usuaria 8; entrevista) |

Conviene destacar que esta misma usuaria (U8) también muestra un comportamiento lingüístico estable en relación con el uso de [“muy” + adjetivo o adverbio], de modo que se puede inferir que el uso de lenguaje evaluativo constituye una característica destacable en la forma de expresarse de esta participante en concreto. Este tipo de perfilamiento que aúna características cuantitativas, midiendo la idiosincrasia de una construcción, como es el MRG, con análisis cualitativos que notan la preferencia de un usuario por cierto tipo de construcciones, es valioso en el perfilamiento de un autor con fines forenses.

Nuestro análisis también arroja luz sobre la estabilidad en el uso de ciertos adjetivos o adverbios que permiten a los usuarios evaluar una situación. Estas construcciones tienen el potencial de constituir marcas lingüísticas individuales, puesto que hay infinidad de recursos disponibles para los usuarios de una lengua a la hora de verbalizar descripciones; así, el hecho de que recurran constantemente al mismo vocablo puede proporcionar información sobre un estilo idiolectal individualizante. Dicho esto, es cierto que cuatro usuarios (44,4 %: U1, U6, U7, U8) usan las mismas construcciones repetidamente, y, más concretamente, el adjetivo “bueno” (MRG = 2,75) y el adverbio “bien” (MRG = 3,63), ejemplificados en (5-6). Así, el uso de estas construcciones no tiene potencial discriminatorio en esta comunidad de habla.

- | | | | |
|-----|--|-----------------------------|---------------------|
| (5) | | el análisis es bueno | (usuario 6; e-mail) |
| (6) | | a esa hora está bien | (usuaria 8; e-mail) |

Empero, la usuaria 2 produce a través de los cuatro géneros de comunicación un adjetivo menos frecuente en comparación con las producciones lingüísticas de los demás usuarios (MRG = 1,67): el adjetivo “complicado”, como medio para evaluar de manera negativa una situación, categóricamente intentando con ella mitigar su calificación, y en ocasiones (7a-b) aunando un intento de empatía con su interlocutor. Así, podemos ver que la frecuencia de uso de una construcción incide en el surgimiento de una función pragmática más especializada

(en este caso, el crear empatía) en contextos específicos, en línea con la gramática de construcciones basadas en el uso.

- (7) a. qué **complicado** (usuaria 2; mensaje de texto)
b. intentemos dar lo mejor para salir fortalecidos de esta **complicada** situación (usuaria 2; *e-mail*)
c. pero la discusión en biología es súper extensa y **complicada**, ¿no? (usuaria 2; entrevista)
d. en algunos casos sabemos que es **complicado** hacerlo (usuaria 2; reunión)

6.2 Expresión de la modalidad

Bybee et al. (1994) afirman acertadamente que puede ser imposible llegar a una caracterización sucinta del dominio nocional de la modalidad, dada la naturaleza de las formas que operan para su manifestación. En el presente artículo, tomamos la noción de *modalidad* como referente a un subcampo semántico del dominio más amplio de las categorías cualitativas, que se ubica junto a dominios como el tiempo y el aspecto (Nuyts, 2016), el cual es el dominio de expresiones de posibilidad (modalidad epistémica) y necesidad (modalidad deóntica) (Van der Auwera, 1996). El alcance de la modalidad es extraproposicional, es decir, presenta alcance sobre toda la cláusula u oración a la que acompaña. En el caso de la lengua española, no existen morfemas modales, y el punto de vista del usuario respecto a lo enunciado puede manifestarse a través de diferentes unidades morfológicas como los adverbios, los verbos cognitivos y modales, las partículas discursivas, etcétera.

La *modalidad epistémica* implica una estimación, generalmente (pero no categóricamente) realizada por el usuario, de la probabilidad de que el estado de las cosas expresado en una cláusula se aplique en el mundo (Nuyts, 2016). También seguimos la definición de Nuyts (2005) en cuanto a *modalidad deóntica*, quien la entiende como una categoría de actitud, donde se indica el grado en el que el asesor (típicamente, pero no necesariamente, el usuario) puede comprometerse con el estado de las cosas en términos de ciertos principios externos a ese estado de las cosas; en este caso, principios morales.

Las secciones 6.2.1 y 6.2.2 entran de lleno en el análisis de la producción de construcciones modales por parte de nuestros participantes a través de los cuatro géneros estudiados. Los resultados ponen el foco en las construcciones modales como marcas lingüísticas determinantes a la hora de mantener la idiosincrasia de la expresión genuina de cada usuario.

6.2.1 Expresión de la modalidad deóntica

La mayoría de los participantes (55,5 %, N = 5: U1, U2, U4, U7, U8) muestra un comportamiento lingüístico estable (RG = 4) a la hora de expresar modalidad deóntica, indicando obligación o necesidad. Cuatro usuarios (44,4 %: U1, U2, U4, U8) exhiben estabilidad en el uso de la construcción [“tener que” + infinitivo], como se muestra en (8).

- (8) a. **tenía que leer** 3 por semana (usuaria 1; reunión)
b. **tendríamos que discutirlo** (usuaria 2; reunión)
c. **tuvimos que reunirnos** al día siguiente (usuaria 4; reunión)
d. **tienes que pensar** cómo unir dos historias (usuaria 8; *e-mail*)

Esta construcción no resulta interesante desde una perspectiva idiolectal, ya que la MRG es de 3,13. Ahora bien, el usuario 7 muestra preferencia por otra construcción de modalidad deóntica: [“haber que” + infinitivo] en enunciados impersonales (RG = 4, MRG = 2,13), lo que sugiere que ha desarrollado un hábito lingüístico poco común y, por consiguiente, potencialmente distinguible de su estilo idiolectal.

- (9) a. por lo pronto, **hay que mantenerse** guardaditos (usuaria 2; mensaje de texto)
b. **habría que preguntarle** a las autoras si están de acuerdo con los cambios (usuario 7; *e-mail*)
c. **hay que complementar** ahí con..., con algo que yo escriba en una *tablet* (usuario 7; entrevista)
d. **hay que actualizar** las fechas (usuario 7; reunión)

Estos resultados muestran que la mayoría de los usuarios en una comunidad de habla se mantienen fieles a las normas lingüísticas de la misma en lo que se refiere a los marcadores deónticos; sin embargo, también hay casos donde uno o pocos usuarios se alejan de la norma, subrayando el potencial de esta área lingüística en los análisis de autoría forense.

6.2.2 Expresión de la modalidad epistémica

Cinco de los participantes (55,6 %: U2, U4, U6, U7, U8) permanecen constantes en relación con ciertas expresiones de modalidad epistémica que les permiten articular la aprehensión subjetiva de la realidad que están expresando. Estas construcciones epistémicas principalmente tienen un efecto en el ámbito discursivo, haciendo hincapié en evaluaciones personales y recalando procesos cognitivos. En primer lugar, cuatro participantes (44,4 %: U2, U4, U6, U7) se sirven de la construcción [“creer que” + cláusula] en los cuatro géneros analizados: hay que notar que *creer* es el más prototípico de los verbos doxásticos o cognitivos (Soler, 2022). Esta construcción no es interesante desde una perspectiva idiolectal, ya que la MRG es de 3,25, y todos los usuarios del corpus la usan por lo menos en un género.

- (10) a. **creo que** eso es todo (usuaria 2)
b. **creo que** es de 8 a 3 (usuaria 4)
c. **creo que** no perdemos nada (usuario 6)
d. **creo que** ahí tenemos un problema (usuario 7)

Sin embargo, hay otras construcciones epistémicas que, como “creer que”, también tienen que ver con la certeza o seguridad de lo que se está diciendo o muestran un valor opinativo: estas son las construcciones [“no sé si” + cláusula] (MRG = 1,75), producidas por la usuaria 4, y [“la verdad es que” + cláusula] (MRG = 2,3), producidas por la usuaria 8. En los usos más comunes, que también se distinguen en los discursos de estas dos usuarias, expresan una opinión, presentando una conclusión personal sobre el tema comentado, donde predomina atenuación de lo dicho, ya que la aserción introducida por “la verdad es que” no describe una verdad extralingüística, sino que sitúa esta en una dimensión cognitiva en la que lo que es verdad no lo es sino para la usuaria (11). Por otro lado, la duda introducida por “no sé si” también intenta mitigar lo dicho, para comprometerse solo mínimamente con ello (12).

- (11) a. **la verdad** no tengo contacto (usuaria 8; mensaje de texto)
b. **la verdad es que** [...] no hay una buena traducción (usuaria 8; *e-mail*)
c. **la verdad es que** en las vacaciones sí- estoy súper cansada (usuaria 8; entrevista)
d. **la verdad es que** en la primera lectura que hicimos [...] se me fue (usuaria 8; reunión)
- (12) a. **no sé si** vaya un 01 antes (usuaria 4; mensaje de texto)
b. **no sé si** entraron los seguros ahí (usuaria 4; *e-mail*)
c. **no sé si** salió una después de esas (usuaria 4; entrevista)
d. **no sé si** tengan alguna pregunta (usuaria 4; reunión)

Algunos usuarios también mantienen un comportamiento lingüístico estable en relación con el uso de partículas discursivas con valor epistémico. En concreto, tres usuarios (33,3 %: U4, U6, U8) emplean el vocablo “bueno” en todos los géneros para crear cohesión en el discurso o señalar el inicio o final de una intervención, así

como de discrepar de su interlocutor. Esta construcción, empleada por ocho de los nueve usuarios del corpus (88,8 %, MRG = 2,6) –aunque no siempre con un RG de 4–, muestra un fuerte arraigo en la comunidad de habla que examinamos.

- (13) a. **bueno**, a mí no me gustan (usuaria 4; entrevista)
b. **bueno**, no importa, acá ya están para abril (usuario 6; chat)
c. **bueno**, piensa qué es lo mejor para ti (usuaria 8; *e-mail*)

Dos de los tres usuarios que usan “bueno” en todos los géneros (22,2 %: U4, U8) también muestran un uso estable de otras construcciones de modalidad epistémica. La usuaria 4 se muestra constante (RG = 4) (a) en el uso de la locución “o sea” (MRG = 1,7) para indicar una clarificación o paráfrasis de lo anteriormente comentado, (b) en el uso de la coetilla interrogativa “¿no?” (MRG = 2,25) para interactuar con sus interlocutores y comprobar su entendimiento o acuerdo con el contenido del mensaje, y (c) en el uso de la locución adverbial “mejor” (MRG = 2,25) para denotar preferencia.

- (14) a. [...] la mamá de mi sobrina, **o sea** mi hermana (usuaria 4; mensaje de texto)
b. ya cubrió 3 vacantes que tenía, **o sea** 14 plazas 29.5 (usuaria 4; *e-mail*)
c. son de- policíacas, **o sea**, cosas como *Mindhunter* (usuaria 4; entrevista)
d. le toca el 5 por ciento de becas, **o sea**, dos becas (usuaria 4; reunión)
- (15) a. en permanentes sí puede haber variación de consejeros, **¿no?** (usuaria 4; mensaje de texto)
b. ya tendríamos que pagar menos, **¿no?** (usuaria 4; *e-mail*)
c. eso era lo que me encantaba, **¿no?** (usuaria 4; entrevista)
d. tuvieron dos semanas para revisar, **¿no?** (usuaria 4; reunión)
- (16) a. como no me bañé el fin de semana no salí y **mejor** vi *Netflix* (usuaria 4; mensaje de texto)
b. **mejor** te paso todo el archivo con las diferencias (usuaria 4; *e-mail*)
c. luego pues ya **mejor** paré (usuaria 4; entrevista)
d. se cambian de carrera o **mejor** se van a hacer otra cosa (usuaria 4; reunión)

Tanto la construcción “o sea” como la construcción “¿no?” fueron producidas por los nueve usuarios del corpus. Sin embargo, solo la usuaria 4 las utiliza en todos los géneros analizados. Dicho esto, y como se puede observar en la tabla 1, la mayoría de los usuarios (N = 6, 66,6 %) tienen un número total de palabras en sus respectivos corpus más alta que la usuaria 4, y aun así es esta usuaria quien hace el mayor (e idiosincrático) uso de estas partículas, dando más fuerza todavía a la hipótesis de que las partículas discursivas pueden mostrar potencial discriminatorio en el campo del análisis de autoría forense.

En esta misma línea, observamos estabilidad en el uso de la locución adverbial “por ejemplo” por la usuaria 8, cuando (a) desea crear empatía con su interlocutor (17a), (b) desea ilustrar lo antes dicho o lo que viene a continuación (17b-c) o (c) desea mitigar y acotar ciertas enunciaciones, como la petición que hace en (17d) o la llamada de atención que hace a sus colegas de trabajo (17e), donde usa “por ejemplo” de manera atenuante en dos ocasiones.

- (17) a. **por ejemplo** (usuaria 8; mensaje de texto)
[en respuesta a un mensaje de su interlocutor, para mostrar que concuerda con él]
b. ahí revisamos el ejercicio 2, **por ejemplo** (usuaria 8; *e-mail*)
c. **por ejemplo**, me encantaba la sopa (usuaria 8; entrevista)

- d. **por ejemplo**, todos esos documentos sería bueno que (usuaria 8; reunión)
nos los trajeran
- e. Yo nada más sí quisiera llamar la atención, **por ejemplo**, en el peligro que hay en ser demasiado
laxos, **por ejemplo**, con el asunto del suplente

El uso doble de “por ejemplo” en una sola enunciación, como vemos en (17e), muestra el arraigo de esta partícula en la mente del usuario, dando fuerza a la noción de que la activación de una unidad cognitiva automatizada a través de la rutina ayuda al lingüista forense a explicar diferencias idiolectales entre autores. La construcción “por ejemplo” la produjeron los nueve usuarios del corpus, aunque solo la usuaria 8 hace uso de la misma en la totalidad de géneros analizados (MRG = 1,9).

Nuestros resultados sobre epistemicidad concuerdan con las conclusiones de Kredens (2002) para el inglés sobre las diferencias idiolectales en la frecuencia de uso de construcciones que ejercen funciones de mitigación, y con las de Litvinova y colegas para el ruso, cuyos resultados apuntan a las partículas discursivas (Litvinova, Litvinova y Seredin, 2018) y a las palabras que describen procesos cognitivos (Litvinova et al., 2018) como rasgos lingüísticos que muestran estabilidad idiolectal.

6.3 Expresiones de cantidad

Los cuantificadores forman una última área de función lingüística recurrente. Más específicamente, las construcciones de cuantificación “mucho”, “un poco” y “cierto” aparecen de manera multigenérica (RG = 4) en la producción lingüística de algunos participantes. A continuación, ofrecemos un análisis de los datos.

Primeramente, cuatro participantes (44,4 %: U2, U4, U6, U8) permanecen constantes al referirse a grandes cantidades, ya sea cuando “mucho” actúa como adjetivo (MRG = 3,13) (18) o como adverbio (MRG = 2,5) (19).

- (18) a. habrá que pensar con **mucho cuidado** qué va a suceder (usuaria 2; e-mails)
- b. usas **muchos adjetivos** siempre (usuaria 8; entrevista)
- (19) a. cuídate **mucho** (usuario 6; mensaje de texto)
- b. eso fue hace **mucho** (usuaria 8; entrevista)

El uso de “mucho” no posee valor idiolectal, dado que parece una preferencia lingüística compartida por la comunidad de habla en general. Empero, la usuaria 8 no solamente se mantiene estable en la expresión de grandes cantidades, sino que también usa la construcción “un poco” de manera idiosincrática (MRG = 1,6) en todos los géneros para atenuar o minimizar lo dicho, encontrándose así esta construcción a medio camino entre cuantificador y marca epistémica (20).

- (20) a. te mando **un poco** más (usuaria 8; mensaje de texto)
- b. dime si con eso te queda **un poco** más clara la novela (usuaria 8; e-mail)
- c. me imagino que pensando **un poco** en la idea de estilo (usuaria 8; entrevista)
- d. se contradice **un poco** (usuaria 8; reunión)

Por otro lado, la usuaria 2 produce la palabra “cierto”, también a caballo entre cuantificador y marca epistémica, en todos los géneros. Esta alta frecuencia multigenérica de “cierto” tiene que ver con la creación de construcciones nominales que establecen gradualidad –es decir, evaluación– en la expresión del compromiso del usuario con lo dicho. Más específicamente, “cierto” atenúa la enunciación al mostrar poco compromiso con lo dicho, puesto que denota imprecisión extensional (Eguren y Sánchez, 2007). La usuaria 2 utiliza “cierto” de una manera estable e idiosincrática (MRG = 1,4), como un cuantificador indefinido, lo que supone un aproximativo evaluativo, cuyo valor es adjetival (21).

- (21) a. Que la gente tenga **cierto** nivel económico no es garantía de que tengan los mínimos de la educación (usuaria 2; mensaje de texto)
- b. Como saben, el fin de semana falleció un profesor de nuestra universidad y esa lamentable pérdida me puso a pensar en torno a la complicada situación que deben estar padeciendo las y los profesores que pasan de **cierta** edad (usuaria 2; *e-mail*)
- c. O sea que concebimos que el- una vida y sobre todo una vida humana tiene **cierta** forma de agencia, tiene formas de relacionar, de aprender, etcétera. (usuaria 2; entrevista)
- d. A mí en lo particular sí me parece que un impulso en el ámbito académico y que en **ciertas** clases de impulso laboral e incluso para poder solicitar una beca, etcétera, tener publicaciones es un gran impulso institucional. (usuaria 2; reunión)

El uso de “un poco” y de “cierto” indican el potencial discriminatorio de las expresiones de cantidad en el campo del análisis de autoría forense. Debemos notar que los cuantificadores que muestran potencial son aquellos que exhiben funciones epistémicas, y, más concretamente, la función de atenuación pragmática, para proteger o reparar la imagen afectada del usuario y/o de sus interlocutores.

7 Conclusiones y direcciones futuras

Este estudio exploratorio constituye una contribución empírica a la noción de idiolecto mediante el análisis de la estabilidad lingüística en el ámbito individual, usando materiales sociolingüísticamente heterogéneos. Nuestros resultados muestran avances significativos respecto a lo que sabemos sobre la estabilidad idiolectal en español, expuestos en los párrafos siguientes.

Primeramente, observamos variación individual dentro de una misma comunidad de habla, donde algunos usuarios producen construcciones potencialmente individualizantes y –de manera más importante– que permanecen estables aun bajo los efectos de género de comunicación (44 %: U2, U4, U7, U8), mientras que otros usuarios producen habla que se conforma a la de la mayoría de los miembros de su comunidad (55,5 %: U1, U3, U5, U6, U9), sin mostrar construcciones idiosincráticas. Estos resultados concuerdan con Mufwene (2010), quien recalca que no hay que creer que las propiedades del habla individual siempre se traducen a normas de la comunidad, puesto que no siempre es el caso.

Se ha visto, además, que la estabilidad idiolectal de algunos usuarios –la cual muestra el uso frecuente de una misma construcción, con un alto grado de peculiaridad, en línea con la metodología necesaria para la resolución de casos forenses (Grant, 2010)– afecta más de un área de función lingüística. Esto es, la mitad de los usuarios que mostraron estabilidad idiolectal con potencial individualizante (U2 y U8) lo hicieron de manera repetida en casi todas las áreas lingüísticas que analizamos, con excepción de la expresión de modalidad deóntica. Dicho de otro modo, las usuarias 2 y 8 se mantuvieron constantes en las áreas de lenguaje evaluativo, de expresión de modalidad epistémica y de cuantificación. Esto apunta a que dentro de la comunidad de habla estudiada hay tres tipos de usuarios: (a) aquellos que en general no muestran un uso idiosincrático de la lengua (U1, U3, U5, U6, U9), (b) aquellos que muestran estabilidad idiolectal en un área lingüística solamente (la U4 en la expresión de modalidad epistémica y el U7 en la expresión de modalidad deóntica), y, finalmente, (c) aquellos que muestran estabilidad idiolectal en varias áreas lingüísticas (U2 y U8), indicando tener un estilo idiolectal muy particular y estable, que aparece en varios contextos y con varios públicos disímiles.

Adentrándonos en el aspecto sociolingüístico de la estabilidad idiolectal, de los nueve usuarios que han sido objeto de nuestro estudio, las construcciones gramaticales que muestran mayor estabilidad idiolectal se concentran en cuatro participantes, tres mujeres y un hombre. Las tres mujeres muestran estabilidad sobre todo en la producción de expresiones epistémicas, mientras que el hombre demostró tener estabilidad en el uso de

marcadores deónticos. Aunque la comunidad de habla estudiada es de tamaño reducido, estos resultados son prometedores; estudios futuros deberían indagar en las diferencias de género y/o sexo en términos de áreas lingüísticas de estabilidad idiolectal.

Nuestros resultados focalizan las áreas lingüísticas más llamativas a la hora de mantener la idiosincrasia de la expresión genuina de cada usuario, en la que ha resultado determinante el área de expresión de modalidad epistémica. En esta área se han distinguido desde verbos hasta sustantivos, adjetivos, adverbios y locuciones adverbiales, partículas del discurso, etc., donde los usos repetidos y particulares de los usuarios generalmente tienden hacia fenómenos de intensidad pragmática, y, más específicamente, hacia la atenuación. Asimismo, en el área lingüística de expresiones de cuantificación, donde también se halla estabilidad idiolectal, son las construcciones a caballo entre la cuantificación y la pragmática las que muestran más potencial a la hora de llevar a cabo análisis de autoría forense.

Cabe destacar que la teoría de la gramática de construcciones basada en el uso es fructífera para poder explicar el uso regular y peculiar de las construcciones idiosincráticas de los participantes del estudio. Es en este trasfondo teórico en el que podemos entender cómo el uso individual de la lengua hace que las construcciones que se arraigan en la mente de un usuario no necesariamente sean las mismas que las que se arraigan en otro, y entender también cómo la frecuencia de uso de una construcción incide en el surgimiento de una función pragmática más especializada en contextos específicos.

Una última conclusión general es que parece que existen sitios de estabilidad individual en términos tipológicos, es decir, áreas funcionales inmunes a los efectos de género a través de diferentes lenguas, que, notablemente, son de diferentes ramas de una familia lingüística. Más específicamente, nuestros resultados en el área de expresión de modalidad epistémica en español concuerdan con resultados en esta misma área para el inglés (Kredens, 2002) y el ruso (Litvinova, Litvinova y Seredin, 2018; Litvinova et al., 2018), donde se halla estabilidad idiolectal.

Una limitación del estudio es el que nuestros datos no estén normalizados, ya que hay un desbalance en el material disponible en relación con los distintos géneros y usuarios –algo esperable partiendo de la naturaleza de los datos, como se aprecia en la tabla 1.⁵ Dicho esto, este es un primer acercamiento al tema con datos naturales, y el hecho de que podamos, por un lado, comparar el uso de determinadas construcciones entre usuarios (*variación interusuario*) a través de la MRG, y, por otro, analizar cuáles son algunas de las áreas lingüísticas y el tipo de construcciones gramaticales del español que sobreviven a los efectos de género de comunicación (*variación intrausuario*), sienta bases más sólidas para el análisis de autoría forense en español, haciendo avanzar este campo.

Es importante comentar que los materiales lingüísticos que recopilamos solamente incluyen las producciones lingüísticas de nueve usuarios. Por tanto, no sabemos si al comparar estos datos con datos de corpus más extensos encontraríamos que las construcciones que aquí hallamos estables seguirían mostrando potencial discriminatorio: estudios futuros deben ocuparse de la necesidad de estudiar una comunidad de habla más amplia, tanto mexicana como de otras variedades del español. Aun así, los casos de análisis de autoría forense no consisten en indicar el autor más probable de un texto entre todos los miembros de una misma comunidad de habla, sino entre los candidatos de una población relevante reducida, la misma situación que propone este trabajo. Este estudio constituye, así, un paso adelante para la creación de una fundamentación teórica necesaria y precisa sobre la estabilidad idiolectal en español en un contexto forense.

⁵ La fórmula para obtener una frecuencia normalizada por cada mil palabras es: [(número de rasgos con un RG de 4 presente en un género / número de palabras por hablante en ese género) * 1000].

8 Referencias

- Anthony, Laurence. (2017). [AntConc](#) (Version 3.5.0) [Computer Software]. Waseda University.
- Baayen, Harald, Van Halteren, Hans, Neijt, Anneke, y Tweedie, Fiona. (2002). An experiment in authorship attribution. *Proceedings of JADT 2002: 6es Journées internationales d'Analyse statistique des Données Textuelles*, 69-75.
- Barlow, Michael. (2013). Individual differences and usage-based grammar. *International Journal of Corpus Linguistics*, 18(4), 443-478. <https://dx.doi.org/10.1075/ijcl.18.4.01bar>
- Bell, Allan. (1984). Language style as audience design. *Language in society*, 13(2), 145-204.
- Blas Arroyo, José Luis. (2017). ¿Identidades lingüísticas invariables? Trayectorias idiolectales ante la variación y el cambio lingüístico en el español clásico. *Revista Internacional de Lingüística Iberoamericana*, 30, 117-144.
- Bloch, Bernard. (1948). A set of postulates for phonemic analysis. *Language*, 24(1), 3-46.
- Bucholtz, Mary, y Hall, Kira. (2010). Locating identity in language. En Carmen Llamas y Dominic Watt (eds.), *Language and identities* (pp. 18-28). Edinburgh University Press.
- Bybee, Joan, y Eddington, David. (2006). A Usage-Based Approach to Spanish Verbs of 'Becoming.' *Language*, 82(2), 323-355.
- Bybee, Joan, Perkins, Revere, y Pagliuca, William. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. The University of Chicago Press.
- Cicres, Jordi, y Queralt, Sheila. (2019). An n-gram based approach to the automatic classification of schoolchildren's writing. *Vigo International Journal of Applied Linguistics*, 16, 53-80.
- Coulthard, Malcolm, Grant, Tim, y Kredens, Krzysztof. (2011). Forensic linguistics. En Ruth Wodak, Barbara Johnstone y Paul E. Kerswill (eds.), *The SAGE Handbook of Sociolinguistics* (pp. 531-544). SAGE Publishing. https://sk.sagepub.com/reference/hdbk_sociolinguistics/n36.xml
- De Houwer, Annick. (2005). Bilingual development in the early years. En *Encyclopedia of Language & Linguistics* (2a ed.). Elsevier.
- Drummond, Rob, y Schlee, Erik. (2016). Identity in variationist sociolinguistics. En *The Routledge handbook of language and identity* (pp. 50-65). Routledge.
- Eckert, Penelope. (2000). *Linguistic variation as social practice*. Blackwell.
- Eguren, Luis, y Sánchez, Cristina. (2007). Grammar and history of Spanish *cierto*. En Georg A. Kaiser y Manuel Leonetti (eds.), *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"*. Arbeitspapier, 122. Fachbereich Sprachwissenschaft, Universität Konstanz, 1-22.
- Gibbons, John, y Turell, Maria Teresa (eds.). (2008). *Dimensions of Forensic Linguistics*. John Benjamins.
- Giles, Howard, y Powesland, Peter. (1975). *Speech style and social evaluation*. Academic Press. <https://dx.doi.org/10.1017/S0047404500005820>
- Goldberg, Adele. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, 7(5), 219-224. [https://doi.org/10.1016/S1364-6613\(03\)00080-9](https://doi.org/10.1016/S1364-6613(03)00080-9)
- Goldstein-Stewart, Jade, Winder, Ransom, y Sabin, Roberta. (2009). [Person identification from text and speech genre samples](#). "Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics", 336-344.

- Grant, Tim. (2010). Text messaging forensics Txt 4n6: Idiolect free authorship analysis? En Malcolm Coulthard y Alison Johnson (eds.), *The Routledge handbook of forensic linguistics* (pp. 536-550). Routledge. <https://doi.org/10.4324/9780203855607.cU33>
- Grant, Tim. (2022). *The Idea of Progress in Forensic Authorship Analysis (Elements in Forensic Linguistics)*. Cambridge University Press.
- Hopper, Paul. (1987). Emergent Grammar. *Berkeley Linguistics Society*, 13, pp. 139-157.
- Iglesias, Olivier. (2016). “Se le quedó mirando”: la atracción de clíticos en un corpus de idiolectos (s. XIX-XXI). En Johannes Kabatek (ed.), *Lingüística de corpus y lingüística histórica iberorrománica* (pp. 424-448). De Gruyter. <https://doi.org/10.1515/9783110462357-019>
- Johnson, Alison, y Wright, David. (2017). Identifying idiolect in forensic authorship attribution: an n-gram textbite approach. *Language and Law = Linguagem e Direito*, 1(1), 37-69.
- Juola, Patrick. (2008). *Authorship Attribution*. Now Publishing. <http://dx.doi.org/10.1561/1500000005>
- Kiesling, Scott F. (2013). Constructing identity. En J. K. Chambers y Natalie Schilling (eds.), *The handbook of language variation and change* (2a ed., pp. 448-467). John Wiley and Sons.
- Kredens, Krzysztof. (2002). Towards a corpus-based methodology of forensic authorship attribution: a comparative study of two idiolects. En Barbara Lewandowska-Tomaszczyk (Ed.), *PALC'01: Practical Applications in Language Corpora* (pp. 405-437). Peter Lang.
- Labov, William. (1972). *Language in the inner city: Studies in the Black English vernacular*. University of Pennsylvania Press. <https://doi.org/10.1177/089124167600400410>
- Labov, William. (1984). Field methods of the project on linguistic change and variation. En John Baugh y Joel Sherzer (eds.), *Language in Use: Readings in Sociolinguistics* (pp. 28-53). Prentice Hall.
- Langacker, Ronald. (1987). *Foundations of cognitive grammar. Volume I: Theoretical prerequisites*. Stanford University Press.
- Litvinova, Tatiana, Seredin, Pavel, Litvinova, Olga, Dankova, Tatiana, y Zagorovskaya, Olga. (2018). On the stability of some idiolectal features. En Alexey Karpov, Oliver Jokisch y Rodmonga Potapova (eds.), *Speech and Computer. SPECOM 2018. Lecture Notes in Computer Science* (vol. 11096). Springer, Cham. https://doi.org/10.1007/978-3-319-99579-3_35
- Litvinova, Tatiana, Litvinova, Olga, y Seredin, Pavel. (2018, 13-16 de noviembre). *Assessing the level of stability of idiolectal features across modes, topics and time of text production* [Ponencia]. 23rd Conference of Open Innovations Association (FRUCT), Boloña, Italia, pp. 223-230.
- Milroy, James. (1992). *Linguistic Variation and Change*. Blackwell.
- Mojedano Batel, Andrea, Soler Bonafont, Amparo, y Kredens, Krzysztof. (En prensa). Epistemic modality constructions as stable idiolectal features: A cross-genre study of Spanish. *International Journal for the Semiotics of Law*.
- Moore, Emma. (2011). Variation and identity. En Warren Maguire y April McMahon (eds.), *Analysing variation in English* (pp. 219-236). Cambridge University Press.
- Mufwene, Salikoko. (2010). SLA and the emergence of creoles. *Studies in Second Language Acquisition*, 32(3), 359-400. <https://doi.org/10.1017/S027226311000001X>
- Nuyts, Jan. (2005). The modal confusion: On terminology and the concepts behind it. En Alex Klinge y Henrik Müller (eds.), *Modality: Studies in form and function* (pp. 5-38). Equinox.
- Nuyts, Jan. (2016). Analyses of the Modal Meanings. En Jan Nuyts y Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood* (pp. 31-49). Oxford University Press.

- Ochs, Elinor. (1992). Indexing gender. En Alessandro Duranti y Charles Goodwin (eds.), *Rethinking context: language as an interactive phenomenon* (pp. 335-358). Cambridge University Press.
- Overdorf, Rebekah, y Greenstadt, Rachel. (2016). Blogs, twitter feeds, and reddit comments: Cross-domain authorship attribution. *Proceedings on Privacy Enhancing Technologies*, 3, 155-171. <https://doi.org/10.1515/popets-2016-0021>
- Queralt, Sheila. (2014). Acerca de la prueba lingüística en atribución de autoría hoy. *Revista de Llengua i Dret*, 62, 35-48.
- Queralt, Sheila, y Giménez García, Roser. (2018). La imitación como contraargumento en peritajes de atribución de autoría: estudio de un caso. *Estudios de Lingüística Aplicada*, 36(68), 131-164.
- Rauniomaa, Mirka. (2003). Stance accretion. Estudio presentado en el *Language, Interaction, and Social Organization Research Focus Group*. University of California, Santa Barbara (febrero de 2003).
- Rickford, John, y McNair-Knox, Faye. (1994). Addressee-and topic-influenced style shift: A quantitative sociolinguistic study. En Douglas Biber y Edward Finnegan (eds.), *Perspectives on Register: Situating Register Variation within Sociolinguistics* (pp. 235-276). Oxford University Press.
- Romano, Manuela, y Porto Requejo, María Dolores. (2013). Emotion, attention and idiolectal variation in radio narratives. *Revista Española de Lingüística Aplicada*, (volumen monográfico), 143-164.
- Solan, Lawrence, y Tiersma, Peter. (2004). Author identification in American courts. *Applied linguistics*, 25(4), 448-465. <https://doi.org/10.1093/applin/25.4.448>
- Soler Bonafont, María Amparo. (2022). *Esto es lo que creo...* El dominio de la opinión en el español hablado a través de los usos construccionales de *creo*. *Círculo de Lingüística Aplicada a la Comunicación*, 91, 219-231. <http://dx.doi.org/10.5209/CLAC.77196>
- Spassova, Maria. (2009). [*El potencial discriminatorio de las secuencias de categorías gramaticales en la atribución forense de autoría de textos en español*](#). [Tesis doctoral]. Universitat Pompeu Fabra.
- Spassova, Maria S., López, Fernanda, Queralt, Sheila, Gavaldà, Núria, Cicres, Jordi, García Barrero, David, y Marquina, Montse. (2014). Idiolectometria forense: estat de la qüestió. En Núria Gavaldà Ferré, Montserrat Forcadell Guinjoan y Raquel Casesnoves-Ferrer (eds.), *Ens queda la paraula: estudis de lingüística aplicada en honor a M. Teresa Turell. Sèrie monografies*, 3. Documenta Universitaria.
- Stamatatos, Efstathios. (2013). On the robustness of authorship attribution based on character n-gram features. *Journal of Law and Policy*, 21(2), 421-439.
- Stamatatos, Efstathios. (2017). Authorship attribution using text distortion. "Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics", vol. 1, *Long Papers*, 1138-1149. <https://doi.org/10.18653/v1/E17-1107>
- Swales, John. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Thompson, Geoff, y Hunston, Susan. (2000). Evaluation: an introduction. En Susan Hunston y Geoff Thompson (eds.), *Evaluation in text: authorial stance and the construction of discourse* (pp. 1-27). Oxford University Press.
- 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.
- Turell, Maria Teresa, y Gavaldà, Núria. (2013). Towards an index of idiolectal similitude (or distance) in forensic authorship analysis. *Journal of Law and Policy*, 21(2), 495-514.

- Troya Déniz, Magnolia. (2015). *Quizá(s) y tal vez* en novelistas de España y América. *Philologica Canariensia*, 21, 109-132. <https://doi.org/10.20420/PhilCan.2015.0036>
- Van der Auwera, Johan. (1996). Modality: the three-layered scalar square. *Journal of Semantics*, 13(3), 181-195.
- 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.
- Wright, David. (2017). Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics*, 22(2), 212-241. <https://doi.org/10.1075/ijcl.22.2.03wri>