

PERITOS LINGÜISTAS

CAZAR A CRIMINALES POR LA LENGUA

Únete a mi grupo de Telegram, ahí encontrarás mis aportes:

<https://rebrand.ly/byneon>

Escanea el código QR



Todos tenemos un **ADN lingüístico**, una huella imborrable que dejamos cuando hablamos o escribimos. Los filólogos Javier Blasco y Cristina Ruiz Urbón utilizan el 'big data' para rastrear estas pruebas, y con sus informes ayudan a resolver casos de plagio, suplantaciones de identidad o denuncias de acoso laboral



Por **BRUNO PARDO PORTO**

Se empieza estudiando a Cervantes y se acaba cazando mentirosos. O criminales. Al menos ese es el extraño camino que han transitado dos profesores de la Universidad de Valladolid, hoy metidos a peritos lingüistas, esto es, gente capaz de adivinar quién está detrás de un texto, ya sea una novela o una amenaza de muerte. O de saber la edad del autor de un anónimo. O su sexo. O lo que estaba sintiendo mientras escribía: ira, miedo, presión. Hay mucho subconsciente volcado en las palabras. Y hay personas capaces de revelarlo.

No son tahúres, sino expertos en datos. En los datos de la lengua.

Se llaman Javier Blasco y Cristina Ruiz Urbón. Él es catedrático de Literatura, experto en el Siglo de Oro y en el Modernismo. Ella está a punto de doctorarse con una tesis dedicada a revisar las atribuciones de obras de Cervantes. Fue en 2008 cuando descubrieron la utilidad de la lingüística forense para las investigaciones filológicas, y con tesón se hicieron expertos en la materia: aprendieron a trabajar con corpus de textos, a medir y cuantificar variables lingüísticas, a programar en lenguaje R, a crear algoritmos. A analizar, al cabo, los cientos de datos que esconde cada frase.

Por ejemplo: calculan la frecuen-



cia con la que pones un nombre específico después de un posesivo, o si antepones la oración principal a la condicional. O si tienes tendencia a escoger palabras de tres sílabas cuando la anterior es de dos. Detectan hasta tus manías fonéticas: a este sonido le suele seguir este otro, etcétera. También miden tu densidad léxica. Tu abanico de adjetivos. Tu fraseo. Tus conectores predilectos. En fin, las posibilidades son casi infinitas.

«Cada persona tiene su idiolecto, un estilo propio a la hora de escribir o de expresarse, que es individual e intransferible», subraya Ruiz Urbón, en su despacho. A su lado, Blasco

asiente: «El dialecto nos diferencia por regiones, el sociolecto por la clase social, el geolecto por la geografía, y el idiolecto es como nuestro ADN lingüístico, una suma de múltiples variables». Esta es la clave del asunto: la huella verbal, que es imposible de borrar, no como la dactilar.

«Dos personas pueden coincidir en el valor de su riqueza léxica, que es una variable, o en el uso de las palabras de función, pero si coinciden en varias empieza a sospechar de que se trata, en realidad, de la misma persona», añade la filóloga.

Esas técnicas, tan novedosas, eran utilísimas para las pesquisas poli-



ciales y judiciales. Así que en 2013, estos docentes, tras unos cuantos trabajos ocasionales, decidieron fundar Agilice Digital, una empresa de la Universidad de Valladolid dedicada a resolver, en principio, conflictos de plagio, pero que ha terminado aceptando encargos muy diversos: desde situaciones de acoso digital hasta suplantaciones de identidad, pasando por la detección de ambigüedades en contratos que confunden al personal o la elaboración de retratos robot a partir de los textos de un desconocido. «Al final usamos la lengua prácticamente para todo. En cualquier conflicto en el que interviene el lenguaje, un lingüista

puede actuar desde un punto de vista científico», asevera Ruiz Urbón.

Ahí va un caso. Una pareja solicita sus servicios porque están recibiendo mensajes anónimos con información delicada sobre ellos. Datos concretos, muy difíciles de saber, muy íntimos. Tienen sus sospechas, así que les entregan los correos electrónicos en cuestión y varios textos de sus sospechosos. Ruiz Urbón y

Blasco hacen sus cálculos, comparan y les dan un nombre: la exmujer del novio, que acaba confesando y pidiendo disculpas.

Y otro más. Un denunciado por pegar a un guardia civil afirma que la declaración de los dos testigos había sido manipulada por el agente de turno, así que acude a ellos. «Ahí lo que se veía es que dos testigos distintos habían declarado práctica-

DE ESTUDIAR A CERVANTES A PARTICIPAR EN JUICIOS

A la izquierda, Javier Blasco y Cristina Ruiz Urbón retratados en la Universidad de Valladolid, donde han desarrollado su empresa de peritajes lingüísticos, Agilice Digital // FRANCISCO J. DE LAS HERAS

mente lo mismo, con puntos y comas. Eso es absolutamente imposible», recuerda Ruiz Urbón, al borde de la risa. «Nunca dos personas distintas reproducen más de cinco palabras seguidas exactamente iguales, salvo que sea un refrán, el nombre de una entidad o una expresión común. A eso lo llamamos 'verbatim'», explica Blasco.

Múltiples variables

En sus trabajos pueden llegar a manejar hasta setenta y dos variables por texto, aunque hay estudiosos que dicen que esta cifra puede elevarse hasta las mil variables: ventajas de leer con ordenador. «En realidad las variables son las mismas de siempre, lo que ocurre es que claro, ponte a medir a mano, sin ayuda de un texto digitalizado, el número de artículos o el de adverbios de una determinada categoría. Sería imposible saber cuál es el porcentaje de veces que aparece una palabra en una novela, por ejemplo. O cuál es el verbo más frecuente en la narrativa de Delibes: es abrir, yo lo sé porque lo he cuantificado», comenta Blasco. «Lo que nos permiten estas tecnologías es la 'distant reading': un trabajo con 'big data', con grandes cantidades de texto. Tradicionalmente estábamos condenados a la 'close reading', detenida en un texto. Eso nos permitía ver el árbol, pero no el bosque», apostilla.

A pesar de este nivel de detalle, precisan ambos, las afirmaciones tajantes no son la norma en su profesión. «Nosotros nunca decimos sí o no. Nosotros trabajamos con escalas de probabilidad verbal», sentencian.

Ahí va otro suceso: un padre acude a ellos para demostrar que los correos que recibía de su hija adolescente, que se había marchado de casa, no eran de ella. Tras comparar los correos sospechosos con otros de los que no había dudas confirmaron que el progenitor estaba en lo cierto. ¿Es imposible engañar al sistema, fingir ser otro? «No. Si eres hábil puedes usar expresiones de otro, sus frases, pero hay muchas otras cosas que se te escapan. O muchas veces te preocupas de imitar al otro y no de quitar lo tuyo. O usas sus palabras pero en distintas posiciones, en distintos contextos», responde Ruiz Urbón.

En ese caso, además, hicieron un perfil lingüístico del suplantador de la hija, que mandaron a la Interpol para ayudar a atraparlo. Estos perfiles los elaboran gra-

Actúan en casos de todo tipo

«USAMOS LA LENGUA PARA TODO. CASI EN CUALQUIER CONFLICTO UN LINGÜISTA PUEDE ACTUAR DESDE UN PUNTO DE VISTA CIENTÍFICO»



Peritos lingüistas // Cazar criminales por la lengua

►►► cias a diccionarios cargados de metadatos, porque hay palabras que se usan a determinada edad o en determinados lugares, y eso se mide, se clasifica. El reto ahora es generar mejores diccionarios, exclusivos para el español y no basados en traducciones del inglés. De hecho, Blasco va a dedicar su año sabático a esto. Lleva ya diez mil palabras registradas. Y quiere llegar a las veinte mil.

«Yo no las defino, yo las clasifico, las preparo para que estos programas puedan funcionar», avisa. Cada palabra, por tanto, lleva una serie de etiquetas: qué clase de palabra es, a qué campo semántico pertenece o incluso si está asociada a sentimientos positivos o negativos. Con esos datos es con los que logran cuantificar la sentimentalidad de un texto, las emociones que en él hay volcadas.

El diccionario que utilizan por ahora tiene unas ocho mil palabras, así que el salto es grande. «Cuando estudiamos un texto encontramos que cerca del 80 por ciento de las palabras están recogidas en el diccionario. Al multiplicar por tres las voces del diccionario esperamos que ese porcentaje ascienda hasta el 90 por ciento... Así podremos sacar más conclusiones. Por ejemplo, en el diccionario actual no existe la categoría de signo político. Nosotros ya hemos conseguido diferenciar las palabras por su carga social, familiar o religiosa. Es todo un reto», remata Blasco.

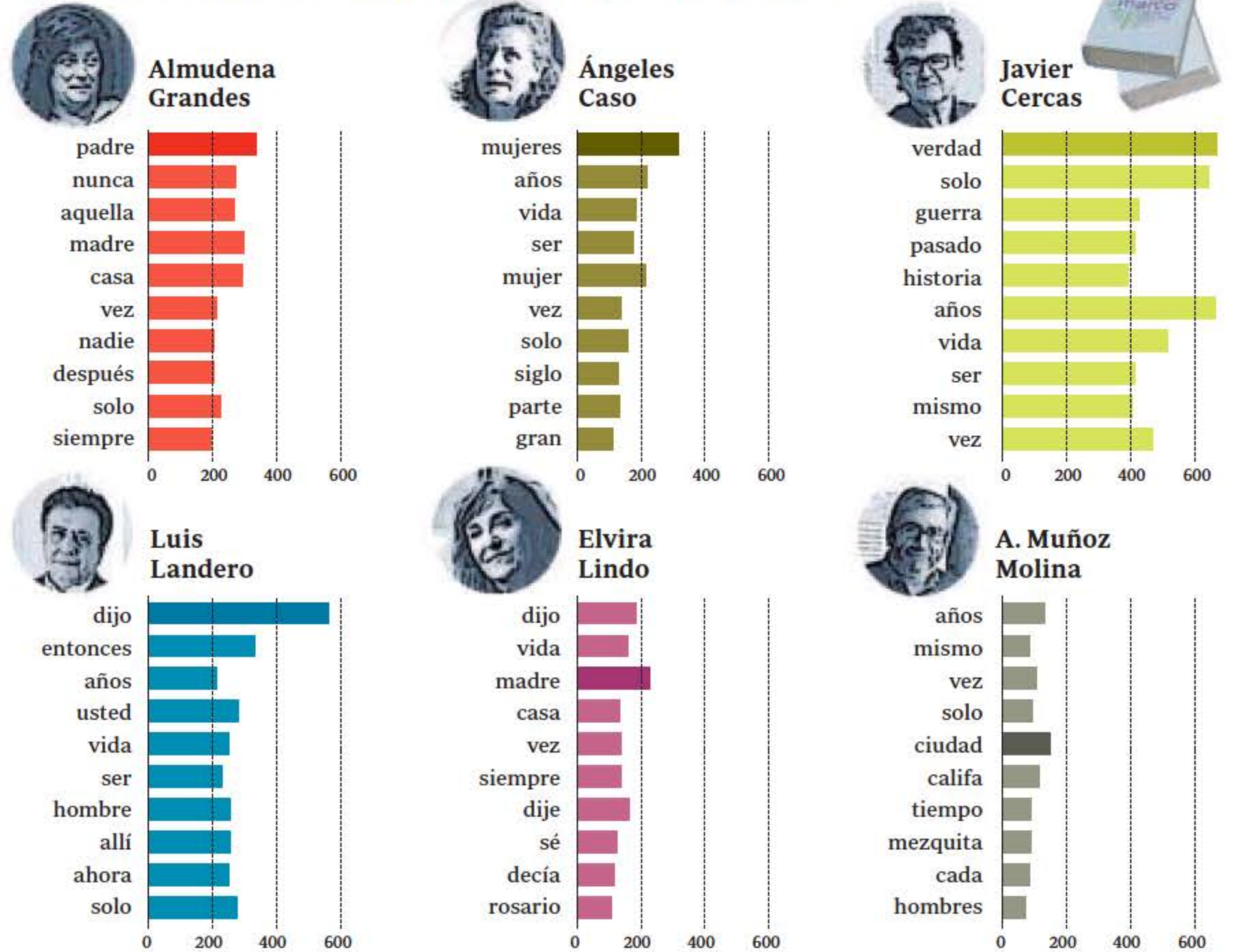
Variedad

Con el paso de los años estos peritos lingüistas se han encontrado con situaciones de todo tipo. Cazaron a un abogado que se había ido a la competencia y estaba incumpliendo el pacto de no concurrencia: estaba participando en casos que involucraban a su anterior empresa, algo totalmente prohibido. También identificaron la mentira de una mujer que decía sufrir una situación de acoso sexual en el trabajo, porque había importantes incoherencias entre las conversaciones de la supuesta víctima con una amiga y con los correos que más tarde envió a su jefe para denunciarlo. Incluso ayudaron a Fernando Trueba a demostrar que el guion de 'La reina de España' era exclusivamente suyo. Por hacer, hasta han corregido a la administración pública más de una vez, en casos de ambigüedad lingüística: la convocatoria de un concurso público que por su redacción no dejaba claro cuándo finalizaba el plazo de solicitud, un cuestionario para detectar falsos autónomos totalmente sesgado...

«Al principio, la gran mayoría de casos que nos llegaban eran de plagio, tal vez por la difusión mediática que se dio a las acusaciones de plagio de escritores como Camilo José Cela, Carlos Fuentes o Bryce Echenique, o a la del ministro alemán Karl Theodor zu Guttenberg,

Las palabras más repetidas por los novelistas

Frecuencia absoluta en una muestra similar en número de palabras de cada autor

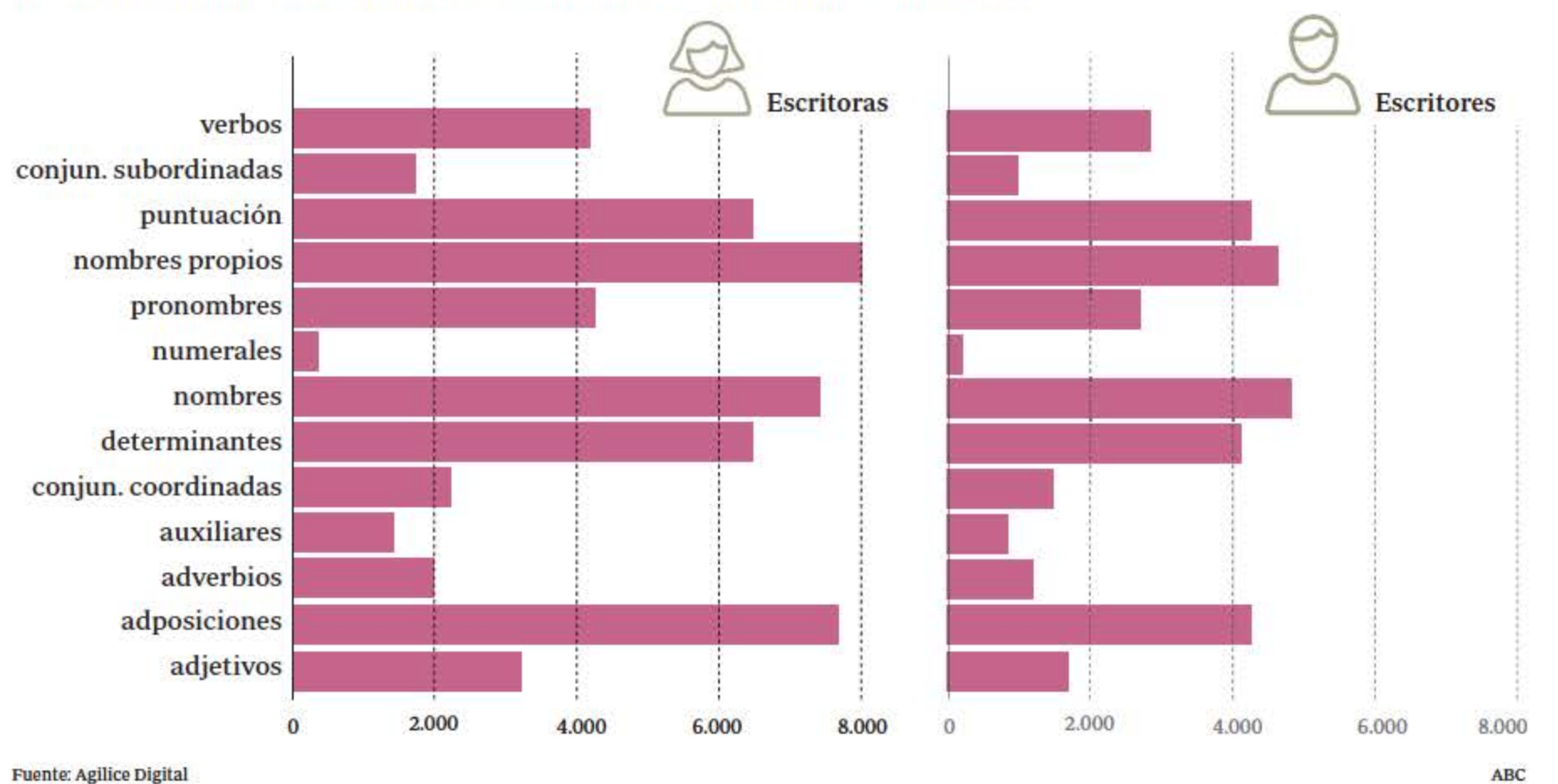


Fuente: Agilice Digital

ABC

Distintas clases de palabras usadas por los autores

Frecuencia absoluta en una muestra similar en número de palabras de cada autor



Fuente: Agilice Digital

ABC

que dimitió de su cargo tras ser acusado de plagiar su tesis doctoral. Por no hablar de otros casos más cercanos. Pero de un tiempo a esta parte la gente está empezando a ser consciente de que usamos el lenguaje para todo, tanto en nuestras relaciones profesionales como personales», repasa Ruiz Urbón. Cada vez los llaman más, y cada vez su labor pare-

LOS NÚMEROS QUE ESCONDE LA LITERATURA

Los peritos pueden saber cuáles son las palabras más utilizadas por los escritores, dibujando su mapa de intereses. También pueden medir las diferencias en la escritura según el género // ABC

ce estar más reconocida. «Las primeras veces íbamos a juicio y decíamos que éramos lingüistas forenses, el propio juez nos miraba como diciendo 'no sé ni lo que es'. Ahora empieza a haber un gran reconocimiento y en muchas de las sentencias se pone en valor nuestro trabajo», celebra Ruiz. Ya se sabe: todo pionero es un Quijote. Al principio.