

ASPECTOS METODOLÓGICOS EN LA CONFECCIÓN DE UN CORPUS JURÍDICO. CONSIDERACIONES A PROPÓSITO DEL *CORPUS DE PROCESOS PENALES*¹

RAQUEL TARANILLA

TRANSLATION AND INTERPRETING INSTITUTE
HAMAD BIN KHALIFA UNIVERSITY (QATAR FOUNDATION)
ragarcia@qf.org.qx

Resumen: Las investigaciones sobre el discurso del derecho suelen recurrir a corpus de textos reales sobre los que fundamentar sus análisis. Ahora bien, el estudio del español jurídico carece todavía de reflexiones profundas sobre cómo deben ser los corpus que emplea. Al hilo de la confección del *Corpus de Procesos Penales*, este artículo plantea algunas cuestiones y presenta algunas propuestas sobre cómo diseñar un corpus para el análisis del español judicial. En concreto, se

describen las necesidades empíricas y las dificultades prácticas de la elaboración del corpus. Asimismo, se da cuenta de las implicaciones éticas de un corpus de textos jurídicos.

Palabras clave: Español jurídico, discurso judicial, lingüística de corpus, ética, anonimización.

Title: Methodological aspects of the construction of legal corpora: Considerations on the Corpus de Procesos Penales.

Abstract: Research on legal discourse tends to

1 La elaboración del CPP fue posible gracias a la colaboración de los Juzgados de lo Penal de la ciudad de Barcelona, que proporcionaron los documentos que lo componen y que prestaron una ayuda inestimable. Una primera versión de este artículo fue presentada como comunicación al IV Congreso Internacional de Lingüística de Corpus (CILC2012), celebrado en la Universidad de Jaén, en marzo de 2012. Su desarrollo se enmarca, parcialmente, en el proyecto *Estrategias de textualidad del discurso profesional en soportes multimodales. Análisis y propuestas de mejora* (FFI2011-28933), financiado por el Ministerio de Economía y Competitividad.

base its analysis in corpora of real texts. However, the study of legal Spanish still lacks theoretical considerations about how its corpora should be. On the basis of the construction of the *Corpus de Procesos Penales* (*Corpus of Criminal Proceedings*), this paper addresses some issues and makes some suggestions on the design of a corpus for

the analysis of judicial Spanish. Specifically, it describes the empirical needs and the practical difficulties of developing corpora. In addition, it deals with the ethical implications of constructing and using a corpus of legal texts.

Key words: Legal Spanish, judicial discourse, corpus linguistics, ethics, anonymization.

1 ESTADO DE LA CUESTIÓN: EL ESTUDIO DEL LENGUAJE JURÍDICO A TRAVÉS DE CORPUS

En la investigación lingüística actual, está plenamente aceptada la idea de que la descripción de una lengua queda enriquecida si se incorpora el análisis de producciones discursivas reales. Basándose en ese fundamento metodológico, la Lingüística de Corpus (Biber, Conrad y Reppen 1998; Hunston 2002; Kennedy 1998; Lüdeling y Kytö 2008; McEnery y Wilson 1996; O’Keeffe y McCarthy 2010; Stubbs 1996; Svartvik 1992; Tognini-Bonelli 2001, entre muchos otros) se ha convertido en una propuesta analítica que ha revolucionado los estudios lingüísticos en pocas décadas (Hunston 2002: 1). Así, la Lingüística ha encontrado en el análisis de corpus una vía preferente para disponer de datos adecuados con los que testar e ilustrar sus hipótesis, así como extraer generalidades, que ha venido a complementar las caracterizaciones de origen especulativo, resultado de la introspección del investigador.

El empleo de corpus de textos reales ha resultado, como era esperable, particularmente atractivo para el estudio de los discursos de especialidad. Surgido en estrecha vinculación con la descripción de textos concretos, el estudio de lenguas en ámbitos específicos encuentra en la metodología de corpus un planteamiento exhaustivo y riguroso sobre cómo organizar, analizar y explotar los textos que constituyen su objeto de análisis (Bowker y Pearson 2002; Flowerdew 2004). En ese sentido, el empleo de un corpus ha pasado a ser poco menos que una garantía de consistencia en los trabajos sobre lengua especializada, por lo que son verdaderamente escasos los estudios de lenguajes específicos que no incluyen un conjunto acotado de textos a partir del cual extraer datos y cuantificar fenómenos.

En concreto, la caracterización del lenguaje propio del derecho también ha experimentado esa promoción de las bondades del empleo de corpus (Tognini-

Bonelli 2001: 1). Por ello, a excepción de los trabajos más generalistas, introductorios o de revisión –como, en el caso del español jurídico, Alcaraz y Hughes (2002), Álvarez (1995), Bordonaba (2009), Samaniego (2004)–, que no suelen estar basados en corpus, las publicaciones actuales en este ámbito acostumbran a incluir una descripción minuciosa del corpus utilizado en el estudio, que actúa, en cierto sentido, como aval empírico de la investigación. Así, contamos, por poner algunos ejemplos, con estudios basados en corpus sobre los géneros de la sentencia (véase, entre otros, Cucatto 2009; López Samaniego 2006a, 2006b, 2010; Taranilla 2009), del informe final que producen abogados y fiscales en el juicio oral (Carranza 2003, 2008), del acta de declaración policial (Taranilla 2006, 2007, 2011a), así como del discurso producido durante el acto del juicio oral (Carranza 2010, Pascual 2003, Ridao 2009a) y en procesos alternativos de resolución de conflictos (Ridao 2007, 2009b).

Ahora bien, como ocurre en los trabajos mencionados, los estudios sobre el discurso producido en contextos jurídicos en español se concentran en un único género, y describen sus características discursivas partiendo de un corpus elaborado específicamente para la investigación particular, que, por esa misma razón, acostumbra a tener un tamaño reducido². La actual atomización del estudio del discurso jurídico en español –que consiste en que se producen trabajos individuales, acerca de géneros discursivos específicos, que no se integran en una propuesta teórica y analítica de conjunto (Taranilla 2011b: 43-44; 2012: 26)– solo podría ser superada si, de una parte, se inicia la reflexión sobre cómo ha de llevarse a cabo el estudio del discurso jurídico a través de corpus y, de otra parte, se elaboran corpus de textos suficientemente representativos y compartidos, que sirvan de base de investigaciones complementarias. Además, puesto que el derecho es un campo en el que se acentúan extremadamente las diferencias entre territorios y culturas, resulta imprescindible que la reflexión sobre la confección de corpus y su empleo como fundamento empírico de los estudios sobre español jurídico mantenga ciertas prevenciones acerca de lo que se viene practicando en otras tradiciones lingüísticas.

2 Salvo algunos corpus multilingües de textos escritos propios de ámbitos especializados, entre los que se incluye el del derecho, tales como el corpus elaborado por el equipo GENTT, *Géneros Textuales para la Traducción*, (<http://www.gentt.uji.es/?q=es/el-corpus>) o por el IULA, *Institut Universitari de Lingüística Aplicada*, (www.iula.upf.edu/corpus/corpus.htm), la realidad es que no contamos todavía con grandes corpus de textos jurídicos en español.

En ese sentido, este artículo pretende, precisamente, dar un paso inicial en la discusión sobre el estudio del español jurídico mediante corpus. Partiendo de la experiencia del diseño y confección del *Corpus de Procesos Penales* (CPP), que fue elaborado con el objetivo de constituir el fundamento empírico de un trabajo sobre la configuración de la narrativa judicial³, este artículo aborda algunas cuestiones relevantes que debe plantearse la investigación en español del derecho a través de corpus.

2 EL DISEÑO DEL CPP (I): ESTABLECIMIENTO DE LAS NECESIDADES

Es un lugar común en la Lingüística de Corpus la idea de que es el propósito de la investigación lo que determina las características del corpus en el que esta se base (véase, por ejemplo, Reppen 2010). Por ello, la pretensión de realizar un estudio sobre cómo se configura el relato de los hechos enjuiciados en sede judicial —que se materializó en los trabajos Taranilla (2011b) y (2012)— requería diseñar un corpus adecuado para tal fin analítico. Este apartado está dedicado a los requisitos que debía cumplir el CPP, que se circunscriben a tres ámbitos: el tema de investigación (§2.1.), las características específicas de la interacción judicial (§2.2.) y el tipo de análisis que se iba a llevar a cabo (§2.3.)⁴.

2.1. Requisitos temáticos

Ningún estudio acerca de una cuestión judicial específica puede orillar que la actividad jurisdiccional no es homogénea, sino que existen distintos órdenes (a saber, penal, civil, social y administrativo) con características particulares. A falta de estudios concretos que permitan realizar afirmaciones rotundas, parece pensable que cada orden jurisdiccional genera productos comunicativos que, en alguna medida, son disímiles. Para ser lo suficientemente precisa y profunda, mi investigación sobre el discurso fáctico en el proceso no podía pretender alcanzar descripciones válidas para todos los actos comunicativos en sede judicial. Era forzoso, pues, limitar

3 El CPP puede encontrarse completo como anexo a Taranilla (2012).

4 Aunque no es el objeto de este trabajo, conviene también tener en cuenta estudios sobre los aspectos formales en la confección de un corpus. Véanse, entre otros, Torruella y Llisterra (1999) y la sección II (“Building and designing a corpus”) de O’Keeffe y McCarthy (2010).

el objeto de análisis y, en consecuencia, restringí el corpus a procesos de una única jurisdicción, la penal. Esa reducción respondía a tres motivos distintos.

Primeramente, la gran mayoría de estudios sobre lenguaje judicial⁵ están realizados sobre corpus de procesos penales⁶, de modo que optar por esa misma opción garantizaba que los resultados de mi estudio fuesen comparables a los obtenidos en investigaciones sobre lenguaje judicial. En segundo lugar, resulta esperable que los juicios penales sean aquellos en los que haya mayor cantidad de secuencias narrativas; por ello, todos los trabajos sobre narrativa judicial utilizan muestras de esta jurisdicción (Carranza 2003; Cotterill 2003; Harris 2001, 2005; Heffer 2005, 2010, entre otros). Como sostiene Gewirtz (1996: 3), el proceso penal es el que mejor sirve a los estudios narrativos, pese a que todo derecho sea, en el fondo, una práctica narrativa. Es más, en el caso concreto del sistema judicial español, hay razones para pensar que en las sentencias penales se dedica mayor atención que en otros órdenes a explicitar cuáles son los hechos que han sido probados durante el juicio oral⁷. En tercer lugar, frente a lo que ocurre en otros órdenes jurisdiccionales, como el civil, donde lo más importante puede llegar a ser el establecimiento de

5 Me refiero a las caracterizaciones sobre el discurso producido en contextos judiciales y dejo al margen los trabajos de corpus en lingüística forense –entendiendo esta disciplina en un sentido estricto, como la participación del lingüista en calidad de perito experto en la fase de prueba de un proceso judicial–.

6 Una de las escasas excepciones es Stygall (1994), que propone un estudio desde el análisis crítico del discurso y la teoría social de inspiración foucaultiana, basado en un único juicio del orden civil. Su preferencia por esta jurisdicción responde a una causa cuantitativa: según Stygall (1994: 4), los juicios civiles son numéricamente muy superiores a los juicios de otros órdenes legales en el sistema de justicia estadounidense. Esto no sería, sin embargo, igual en el sistema español, en el que, según la *Memoria Anual 2011* publicada por el Consejo General del Poder Judicial, la jurisdicción penal ingresó, en el año 2010, 6.639.356 asuntos y emitió un total de 725.281 sentencias, mientras que la jurisdicción civil ingresó, en ese mismo año, 1.984.098 asuntos y dictó 498.847 sentencias.

7 Según han considerado algunos autores (véase, por ejemplo, Montero Aroca 2000: 423-434), mientras que en los órdenes penal y laboral la declaración de hechos probados en la sentencia es un imperativo legal, en los órdenes civil y administrativo, en cambio, no resulta imprescindible que en la sentencia se determinen explícitamente los hechos que fueron probados durante el juicio oral. En efecto, todavía hay quienes consideran que la redacción del artículo 209.2 Ley de Enjuiciamiento Civil (LEC) fundamenta la optatividad de los antecedentes de hecho de la sentencia civil en el redactado “se consignarán [...] los hechos probados, en su caso” (cfr. Igartua 2003: 8 n4). No faltan, sin embargo, autores que consideran que en todos los órdenes jurisdiccionales debe existir la obligación de determinar qué hechos se consideran probados. Al margen de polémicas sobre lo que sucede en la práctica procesal y cómo deberían ser las cosas, es un hecho que el proceso penal contiene efectivamente la declaración de los hechos probados; por ello, en el diseño del CPP se optó por atender a este orden.

acuerdos entre las partes⁸, los juicios en el orden penal tienen el propósito básico de buscar la verdad; por esa razón, los mecanismos de construcción narrativa deberían tener, al menos en teoría, mayor presencia y trascendencia en un corpus penal que en uno del orden civil, donde parece esperable que primen otros mecanismos, como los de negociación y obtención de compromisos.

Al margen del orden jurisdiccional escogido, no bastaba únicamente con que el corpus se limitase a procesos penales, sino que convenía circunscribirlo a la primera instancia jurisdiccional; es decir, se prescindió de los eventos judiciales que tengan lugar en la fase de recurso. Ello se justifica porque es en la primera instancia judicial donde se concentra, salvo excepciones muy puntuales, el grueso de la práctica de la prueba. Asimismo, se prefirió confeccionar un corpus de procesos seguidos por la vía del procedimiento abreviado⁹, y rechazar los procesos que se siguen por la vía del procedimiento ordinario (esto es, los que se siguen para enjuiciar delitos que llevan aparejada una pena privativa de libertad superior a 9 años), debido a que, como han puesto de manifiesto algunos autores (Ferreiro 2003: 3, Garberí 2000: 76, Moreno Catena et al. 2000: 2541), el procedimiento abreviado constituye el auténtico proceso común del sistema de enjuiciamiento español y la vía por la que se tramitan la mayor parte de procesos del orden penal. En ese sentido, resulta esperable que las características que se identifiquen en procesos abreviados estén presentes también en los de tipo ordinario. Además, en relación con cuestiones de tipo práctico, intentar dar cuenta de los procesos narrativos globales de un proceso a partir del análisis de juicios demasiado largos y complejos exige una tarea extremadamente costosa (Harris 2001: 72)¹⁰.

8 Si bien esta concepción negocial de la verdad en el proceso civil es duramente criticada por algunos autores (véase, como muestra, Taruffo 2008), lo cierto es que esos mismos autores reconocen que esta ideología sobre la prueba civil tiene calado en la práctica (Taruffo 2008: 139-140). Por ello, el diseño del CPP prescindió del proceso civil y puso atención en el orden penal, en el que, pese a existir múltiples desacuerdos teóricos acerca de la concepción de la verdad, parece una cuestión pacífica el rol privilegiado que desempeña la búsqueda de la verdad.

9 Para una caracterización detallada de este procedimiento, véase Muerza (2002).

10 Con vistas a la confección práctica del corpus, optar por el procedimiento abreviado tiene la ventaja de que permite dirigirse a los Juzgados de lo Penal (que se encargan de procedimientos abreviados en los casos en los que la pena de privación de libertad solicitada por la acusación es menor a 5 años), sin necesidad de tener que poner muchas condiciones a los documentos que se demandan. Dirigirse, en cambio, a otro tipo de juzgados supone complicar notablemente la solicitud de colaboración a las autoridades jurisdiccionales: así, por ejemplo, la Audiencia Provincial, que conoce de los procedimientos abreviados en los que se solicita pena de prisión de más de 5 años, conoce también de los procedimientos ordinarios. Esa simplificación resulta conveniente ya que uno de los objetivos del diseño del corpus era facilitar la colaboración de los órganos de justicia, de cara a que se decidiesen a contribuir con este trabajo.

Igualmente, se desestimó el estudio de juicios celebrados ante un tribunal de jurado popular, puesto que en este tipo de procesos es esperable encontrar ciertos recursos comunicativos que no resultan de interés en este estudio: por ejemplo, hay argumentos para pensar que los recursos de divulgación del conocimiento técnico, así como ciertos mecanismos de persuasión ligados a las emociones serán más productivos y, por tanto, más frecuentes en procesos judiciales en los que la decisión es tomada por un grupo de ciudadanos legos en derecho (cfr. Casanovas 1997, Storey-White 1997).

2.2. Requisitos discursivos

Por lo común, los estudios sobre discurso narrativo judicial y, en general, sobre el discurso generado en sede judicial han sido realizados en inglés para juicios celebrados en los Estados Unidos o el Reino Unido, de modo que se ha privilegiado el estudio de los eventos comunicativos que tienen una función más destacada en esos países: las instrucciones y, en realidad, todos los actos de comunicación con el jurado popular (Heffer 2005, Tiersma 2006, Vidmar y Hans 2007)¹¹; la celebración del juicio oral (Cotterill 2002, 2003; Heffer 2005), con especial acento en los interrogatorios (Drew 1985, 1992; Philips 1987, entre muchos otros) y el informe final de las partes (Felton Rosulek 2008, 2010; Walter 1988).

Sin embargo, como se ha apuntado arriba, la confección de un corpus de textos judiciales no puede obviar las peculiaridades procesales de cada sistema jurisdiccional. En el sistema de enjuiciamiento español, como ocurre en el resto de los países de derecho continental, la administración de justicia y los actos comunicativos que en ella se producen tienen unas características propias que hay que considerar. De ese modo, si mi investigación final pretendía examinar cómo se configuran los relatos de los hechos en el seno del proceso judicial español, el CPP debía contar con todos aquellos textos que participan en la generación de información factual.

11 A diferencia de lo que ocurre en otros países, el número de procesos con jurado popular en España es limitado; así, según cifras emitidas por el Consejo General del Poder Judicial en su *Memoria Anual 2011*, el número de sentencias dictadas en procesos con jurado popular ha sido de 312 en 2010 (de un cómputo total de 725.281 sentencias en todo el orden penal), sin que ese dato parezca que vaya a crecer en los próximos años: algunos autores hablan, de hecho, de una progresiva "huida del jurado" (Gómez Colomer 2001: 303), intencional por parte de los operadores jurídicos.

Por todo ello, el CPP fue diseñado desde el ánimo de incluir los géneros que, en mayor medida, están implicados en la configuración narrativa de los hechos enjuiciados, esto es, el *sistema de géneros*¹² en el que se configuran y se relacionan los relatos fácticos de un proceso. En efecto, restringir el análisis de la construcción narrativa al evento del juicio oral devuelve al investigador una imagen limitada y superficial de la comunicación judicial. Por ese motivo, el CPP debía partir, al menos, de los escritos de la acusación y la defensa, que contienen la narración de los hechos que sostiene cada parte. Asimismo, el CPP tenía que integrar el juicio oral, que es el evento donde el juez accede directamente a las versiones de los hechos confrontadas, fundamentalmente, a través de los géneros del interrogatorio y la toma de declaración a los peritos. Por último, era imprescindible considerar la sentencia judicial, como último texto del sistema, que establece cuál es la versión de los hechos que se considera probada. Incorporando esos géneros judiciales, el CPP se aseguraba poder comprender el proceso de creación del relato en toda su dimensión dinámica y sistémica. Asimismo, aunarlos en un único corpus de estudio daba respuesta a la idea creciente en el estudio de los discursos profesionales de que la producción y la gestión de la información en estos ámbitos no pueden concebirse sino como actividades multimodales, en las que se combinan y suceden géneros orales y escritos (Maley 1994), que, desde luego, no pueden actuar ni ser comprendidos de forma independiente.

De todos modos, el sistema de géneros del proceso penal en España no se limita al escrito de las partes, la celebración del juicio y la sentencia posterior. Es innegable que en la etapa de instrucción, que es la etapa de investigación previa al juicio, se generan una serie de textos (orales, como entrevistas e interrogatorios judiciales y policiales, y escritos, como informes diversos), agrupados en el *sumario* —o las *diligencias previas*, que es el nombre que recibe el sumario cuando el procedimiento es abreviado—, que de forma evidente participan en la configuración de las narrativas. Así, por ejemplo, el acta de declaración policial es un género que desempeña un papel de primer orden en la construcción del relato de los hechos¹³.

12 La idea de *sistema de géneros* (Bazerman 1994) hace referencia a la existencia de géneros interdependientes, que ocurren en determinadas secuencias típicas, y cuyo propósito y forma están en estrecha interacción.

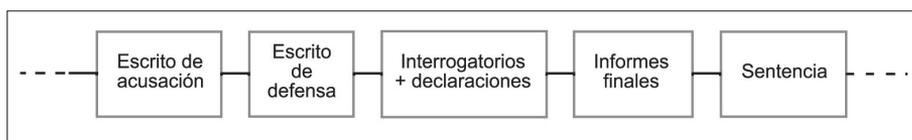
13 Véase, en relación a la dimensión narrativa del acta de declaración policial, Taranilla (2007 y 2011a)

Sin embargo, en el diseño del CPP se decidió prescindir de cualquier documento compuesto durante la fase de instrucción por dos razones.

En primer lugar, la disponibilidad de los textos de la instrucción es necesariamente parcial: si bien se puede solicitar el acceso a los documentos escritos que conforman el sumario, el hecho es que no hay manera de poder disponer de los textos orales que tienen lugar en ese momento del proceso, salvo que al investigador se le permitiese presenciarlos y grabarlos, en una tarea de tipo etnográfico, dado que las labores de investigación no se registran en soporte audiovisual. En segundo lugar, está justificado centrar la atención en los textos que tienen una presencia directa en la fase de juicio oral debido al principio de inmediación que rige el sistema de enjuiciamiento español: se trata de la garantía de que las partes del proceso interactúen, contraponiendo sus versiones de los hechos y las pruebas que las sustentan, de forma no mediada ante el juez que ha de emitir el veredicto. Ese principio, como señala Andrés Ibáñez (2003: 58), implica la prohibición de introducir en el acto del juicio cualquier material extraído directamente de la instrucción, es decir, todo elemento fáctico o jurídico que considere el juez en su decisión ha de resultar de la práctica realizada durante el juicio oral. En ese sentido, el aislamiento preceptivo del juicio oral, que obliga a que el relato de cada una de las partes sea por entero planteado y negociado en el seno del juicio, justifica la decisión de dejar al margen del CPP todos aquellos documentos que no tengan presencia inmediata en el acto del juicio.

En suma, el CPP debía estar integrado (i) por los escritos de las partes, que contienen las versiones de los hechos que, tras la investigación, mantiene cada parte procesal, (ii) por el evento del juicio oral (que incluye los géneros del interrogatorio y la declaración pericial, así como los informes finales de las partes) y (iii) por la sentencia, que establece cuáles son los hechos que se han probado efectivamente. Ello está representado en la figura 1:

FIGURA 1. EL SISTEMA DE GÉNEROS PANALES QUE INTEGRAN EL CPP
(Extraído de Taranilla 2011b)



Con todo, aunque no son consideradas en el diseño del CPP, cabe mencionar que las versiones distintas acerca de los hechos enjuiciados no solamente son contenidas en géneros judiciales (esto es, los producidos en contextos judiciales, por los abogados, el fiscal, la policía, el juez o el secretario, en conjunción con las personas implicadas en la causa), elaborados tanto en la instrucción como en la fase de juicio oral. En realidad, existen otros géneros que pueden concurrir en la configuración narrativa de los hechos enjuiciados¹⁴. Así, por ejemplo, ocurre en los procesos judiciales que tienen trascendencia mediática, en los que los medios de comunicación de masas proponen relatos propios que conviven con los que producen los operadores jurídicos (Latorre 2002). No obstante, por lo que a la investigación en narrativa judicial se refiere, carecen de interés los juicios con especial repercusión en la prensa –tales como el juicio por los atentados del 11 de marzo de 2004 en Madrid, los procesos más mediáticos contra miembros de la banda terrorista ETA, procesos por delitos de corrupción o los juicios por cualquiera de los crímenes que llenan las páginas de crónica social–. Ello es debido, en primer lugar, a que en esos juicios concurren materiales informativos de origen no judicial y, en segundo lugar, a que en ellos a menudo intervienen factores (políticos, sociales, etc.) difíciles de valorar.

2.3. Requisitos analíticos

En cualquier investigación basada en corpus, una de las cuestiones cruciales es la relativa al tamaño que debe tener el corpus, lo que está en estrecha relación con el papel que este desempeñe en la investigación. Frente a los estudios *corpus-driven*, que emplean grandes corpus para obtener datos a partir de los cuales poder inducir teorías generales y categorías nuevas, la investigación que originó el CPP se proponía un estudio basado en *corpus* (*corpus-based*¹⁵), que emplease

14 En este punto, la obra de Foucault (1973) *Moi, Pierre Rivière* es un ejemplo ilustrativo del universo de géneros de variada naturaleza que se reúnen en torno a un hecho, configurando los diversos relatos. Con el propósito de observar cómo el artefacto cultural se apropia y asimila un hecho determinado, Foucault compila informes médicos, textos judiciales e incluso parte de la memoria personal redactada por *Rivière*, acusado de matar a su madre y a sus dos hermanos.

15 Para una descripción detallada de los estudios *corpus-based* frente a los estudios *corpus-driven*, puede verse, por ejemplo, Tognini-Bonelli (2001) y McEnery, Xiao y Tono (2006).

una compilación de textos reales para corroborar o desestimar hipótesis de partida y proporcionar ejemplos de los productos discursivos que iban a ser caracterizados¹⁶; en otras palabras, mi investigación iba a recurrir al corpus como medio para testar e ilustrar las teorías y descripciones que resultan de la labor especulativa del investigador (Tognini-Bonelli 2001: 65). Por ese motivo, el CPP no requería tener un tamaño demasiado grande: en realidad, un corpus relativamente pequeño parecía suficiente para caracterizar la dinámica de la configuración del relato fáctico en el proceso penal¹⁷.

En efecto, varios de los inconvenientes que el Análisis del Discurso encuentra en el estudio de corpus (véase al respecto Baker 2006; Biber, Connor y Upton 2007; Flowerdew 2008: 13-20; Hunston 2002; Virtanen 2008)¹⁸ pueden ser superados si se emplean corpus reducidos. Fundamentalmente, los grandes corpus tienen el peligro de proporcionar segmentos textuales aislados y descontextualizados, de modo que se pierden dos de los elementos, la estructura del texto y su contexto, que mayor significación tienen para los estudios del discurso; en cambio, el uso de un corpus de menor tamaño permite al estudioso conocer con profundidad

16 Esta forma de análisis de corpus hecha fundamentalmente a mano –aunque se recurra a algunas herramientas informáticas para acelerar ciertas búsquedas– seguirá siendo por el momento, según Lee (2008: 89), la más habitual en Análisis del Discurso.

17 Tal como apuntan Bhatia, Langton y Lung (2004: 207), en el lenguaje jurídico, donde las relaciones forma-función son a menudo formulaicas y la configuración textual es tan conservadora, en ocasiones no son necesarios grandes corpus, sino que basta el análisis exhaustivo y adecuado de varios textos, sin instrumentos informáticos de la Lingüística de Corpus. Esta peculiaridad metodológica del estudio del discurso del Derecho, bajo mi punto de vista, puede describirse como la rápida *saturación* (Bowen 2008) de algunos corpus jurídicos: la saturación de un corpus se produce cuando la incorporación de nuevos elementos ya no arroja ni nuevos datos (saturación de datos) ni nuevos aspectos teóricos (saturación teórica) (Bowen 2008: 140); debido a que, en muchas ocasiones, los textos jurídicos guardan entre sí grandes similitudes, los corpus formados por este tipo de textos se saturan enseguida. Por otra parte, en relación a la composición de corpus especializados, Koester (2010: 66) también apunta que, sin restar importancia a las desventajas evidentes de manejar corpus pequeños, lo cierto es que pueden ser suficientes para determinados análisis; véase, también en este sentido, Sinclair (2005).

18 Sobre las diferencias metodológicas que guían la Lingüística de Corpus, por un lado, y el Análisis del Discurso, por otro, véanse las palabras de McEnery, Xiao y Tono (2006: 111): “while DA [Discourse Analysis] emphasizes the integrity of text, corpus linguistics tends to use representative samples; while DA is primarily qualitative, corpus linguistics is essentially quantitative; while DA focuses on the contents expressed by language, corpus linguistics is interested in language per se; while the collector, transcriber and analyst are often the same person in DA, this is rarely the case in corpus linguistics”.

tanto el texto completo como su contexto (Koester 2010: 67). La familiarización del investigador con el objeto de estudio todavía es mayor si el corpus se crea *ad hoc* para el trabajo concreto, buscando, seleccionando, compilando y transcribiendo, si fuese necesario, los documentos escogidos (Baker 2006: 25). De hecho, una parte considerable de los estudios sobre análisis del discurso en sede judicial han recurrido al estudio de un único caso que, bien comprendido por el analista, ha permitido hacer aportaciones relevantes a la disciplina: es el caso, entre otros, de Cotterill (2003), que examina el juicio por homicidio contra O.J. Simpson; de Jacquemet (1996), sobre el macro-juicio italiano contra la Camorra napolitana; de Harris (2005), sobre el juicio por agresión sexual contra el locutor deportivo Marv Albert; o de Hiltunen (1996, 2004), acerca de los juicios por brujería de Salem.

3 EL DISEÑO DEL CPP (II): LAS LIMITACIONES DE LA CONFECIÓN DEL CORPUS

Como afirma Hunston (2008: 156-157), “[a]ll corpora are a compromise between what is desirable, that is, what the corpus designer has planned, and what is possible”. Así, las aspiraciones empíricas que se han descrito en el apartado anterior (esto es, obtener los escritos de las partes, los juicios orales y las sentencias de un pequeño número de procesos penales) cuentan inevitablemente con algunas limitaciones prácticas.

Es un hecho que las relaciones de colaboración entre la universidad y los órganos jurisdiccionales son escasas¹⁹ y, por tanto, que un investigador universitario

19 Entre las excepciones, puede citarse el IDT (Institute of Law and Technology), dirigido por el Dr. Pompeu Casanovas, de la Universitat Autònoma de Barcelona (<http://idt.uab.cat/>), que ha llevado a cabo investigaciones en las que se han involucrado diferentes entidades públicas. Asimismo, es reseñable que, como resultado del Pacto de Estado para la Reforma de la Justicia, se creó la Comisión para la Modernización del Lenguaje Jurídico, por orden JUS/3126/2003, de 30 de octubre; dicha Comisión cuenta con el asesoramiento de cuatro grupos de investigadores en Lingüística española: uno, dedicado al estudio de la lengua judicial oral, dirigido por el Dr. Antonio Briz (Universitat de València); el segundo, que se ocupa de la escritura judicial, dirigido por la Dra. Estrella Montolío (Universitat de Barcelona); el tercero, que se encarga del estudio de los formularios, dirigido por el Dr. Julio Borrego (Universidad de Salamanca); y, el cuarto, encargado del análisis de los manuales de redacción jurídica, dirigido por el Dr. Salvador Gutiérrez (Universidad de León).

obtenga documentos reales generados por la administración de justicia suele ser un proceso plagado de dificultades tanto en España como en otros estados de nuestro entorno²⁰. En España, al margen de bases de datos públicas, como el fondo del Centro de Documentación Judicial (CENDOJ), que es un órgano técnico dependiente del Consejo General del Poder Judicial; o de bases privadas, como WestLaw, que permiten el acceso general a resoluciones judiciales, lo cierto es que todavía no disponemos ni de un corpus de textos jurídicos exhaustivo que pueda considerarse suficientemente representativo de los géneros jurídicos en español, ni de un protocolo de colaboración y acceso de los investigadores universitarios a la actividad jurisdiccional española²¹. De ese modo, en muchos casos, acaba siendo decisión unilateral de alguno de los juristas implicados en un proceso que el investigador pueda, finalmente, tener acceso a textos reales²².

Los motivos que explican la dificultad de conseguir textos pueden ser variados. De un lado, la administración de justicia ha gozado de un estatus privilegiado en comparación con otras instituciones públicas; es obvia su aún escasa transparencia y el limitado conocimiento que los ciudadanos tienen de la actividad real en los tribunales de justicia, a pesar de que haya elementos correctores, como la publicidad de las vistas y las sentencias, o la institución del jurado popular. De otro lado, resulta innegable que la administración de justicia –y, muy particularmente, los procesos del orden penal– maneja situaciones en las que se ven implicadas personas anónimas, de modo que parecería poco decoroso

20 Así, por ejemplo, ocurre en el Reino Unido: en su investigación sobre narrativas judiciales, Heffer (2005: xvii) lamenta lo complicado de acceder a los juicios de las Crown Courts, donde no se puede realizar grabaciones; ello le obliga a trabajar con transcripciones hechas por trabajadores del juzgado (que no son lingüistas), que generalmente son incompletas (Heffer 2005: 52-54).

21 En español, trabajan con corpus reales el grupo de investigación *Grupo de Estudios Sociolegales*, integrante del IDT (nota 18), dedicado al análisis etnográfico (véase Poblet 1998); en concreto, pueden consultarse Cubells (2002) y Pascual (2003). Asimismo, véanse Carranza (2003, 2008, 2010) y Ridao (2007).

22 Así, según mi experiencia, el estudioso de los eventos comunicativos jurídicos a menudo acaba consiguiendo textos reales, de un modo un tanto furtivo, cuando algún profesional del derecho, o incluso alguna de las personas implicadas, accede a darle copia de algunos documentos. Esta situación, sin embargo, hace difícil que el investigador pueda configurar un corpus solvente, representativo y equilibrado, si todos los textos son proporcionados por el mismo agente jurídico (así, por ejemplo, no se podrían extraer conclusiones generales sobre el género del contrato en español a partir de un corpus, aunque fuese grande, compuesto por textos redactados por uno o pocos abogados).

un sistema judicial completamente abierto y exhibido. Puesto que el acceso a documentos escritos y orales implica tener conocimiento de datos e informaciones de carácter personal, se impone considerar las implicaciones éticas de este tipo de investigación.

Así, ante el interés de utilizar grabaciones de juicios orales reales, se hace imprescindible plantear cómo tratar la imagen y los datos de los individuos que intervienen en tales eventos. De un lado, es cierto que el juicio oral es un momento procesal que debe ser, salvo en ciertos casos, público, de modo que cualquier ciudadano pueda asistir a su celebración; sin embargo, un investigador no puede ampararse en el principio de publicidad para utilizar las grabaciones de los juicios indiscriminadamente, sin contar con la autorización de las personas que aparecen en las imágenes o de las que se habla durante el juicio. De ese modo, igual que hay algunos trabajos que, desde la etnografía, plantean las implicaciones éticas de las grabaciones no autorizadas —por ejemplo, Harvey (1992), sobre grabaciones clandestinas en comunidades peruanas—, en el ámbito de los discursos profesionales se echan de menos estudios centrados en el desarrollo ético de la investigación. Una de las escasas referencias sobre esta cuestión es Miller (2000), que reflexiona, desde la óptica del abogado, acerca de las implicaciones éticas de contar las historias de los clientes en artículos de investigación o en clases para futuros abogados. Para Miller, resulta problemático considerar que solo cambiando los nombres, los datos y algunos detalles de las historias se pueda utilizar la historia de alguien libremente. Este autor plantea la necesidad de que se consulte a los protagonistas de la historia, para que puedan opinar sobre el uso de su historia. Sin embargo, esa opción resulta impracticable en la investigación sobre discurso judicial, puesto que resultaría demasiado complejo obtener el permiso de aquellos que en su día participaron en los procesos de los que se derivan los documentos empleados. Por ello, podría obviarse el requisito de la autorización individual, a cambio de poner el máximo esmero en la preservación de su identidad.

En vista de todo ello, en la confección del CPP se tomaron dos medidas de cara a garantizar la ética de la empresa: la firma de un compromiso ético y la anonimización de los datos. De esas dos medidas, así como del proceso general de recolección de datos y confección del corpus, se da cuenta detallada en el apartado siguiente.

4 EL PROCESO DE CONFECCIÓN DEL CPP

4.1. Primera etapa: la obtención de los documentos

Para acceder a los documentos que conforman el CPP, solicité colaboración al Juzgado Decano de Barcelona. En concreto, se solicitaron tres tipos de documentos relativos a veinte juicios penales: (i) los escritos de las partes y del ministerio fiscal, (ii) las grabaciones audiovisuales del acto del juicio oral²³ y (iii) la correspondiente sentencia. Era importante que los procedimientos seleccionados no finalizasen en conformidad del imputado (esto es, con su aceptación de la pena que la acusación solicita para él, que hace innecesaria la celebración del juicio), ya que en ese caso se pierde la confrontación de relatos.

Al recibir la solicitud, el juez decano delegado de los juzgados de lo penal de Barcelona realizó un llamamiento al resto de los juzgados penales de la ciudad, animándolos a colaborar con este estudio. En última instancia, fue en el seno de cada juzgado (esto es, entre el juez titular y el secretario judicial) donde se decidió si contribuir o no con esta investigación.

Teniendo muy presentes las implicaciones éticas expuestas, redacté un compromiso ético de la investigación²⁴, en el que se asumían algunas obligaciones

23 La entrada en vigor de la LEC 1/2000 introdujo la obligación de que, en el orden civil, los juzgados y tribunales españoles registrasen audiovisualmente las actuaciones orales (Gómez Martínez 2003). Posteriormente, la reforma de la Ley de Enjuiciamiento Criminal (LECrim) extendió la misma innovación al orden penal (Magro 2010); así, el nuevo artículo 743.1 de la LECrim reza lo siguiente: “El desarrollo de las sesiones del juicio oral se registrará en soporte apto para la grabación y reproducción del sonido y de la imagen. El Secretario judicial deberá custodiar el documento electrónico que sirva de soporte a la grabación. Las partes podrán pedir, a su costa, copia de las grabaciones originales”. Esas grabaciones que realizan los juzgados fueron las que se emplearon en la confección del CPP, que, por lo tanto, no están realizadas ex profeso para la investigación lingüística. Ello comporta una ventaja: que las personas que intervienen actúan de forma natural, sin que la presencia de un observador –y de su cámara registradora– les condicione; pero también tiene una desventaja: que la cámara fija en el techo, más o menos a la altura de la sala en la que se encuentra el tribunal, grabando en plano general la sala y en una calidad irregular de audio y, sobre todo, de imagen, impide que la investigación aborde algunos aspectos comunicativos, como los referentes a la comunicación no verbal.

24 Para la elaboración de este documento recibí asesoramiento del Comité de Bioética de la Universitat de Barcelona, si bien esta institución todavía no dispone de un protocolo para investigaciones en ciencias sociales y humanas con implicaciones éticas.

referentes al tratamiento del corpus, a su conservación y a la difusión de los resultados obtenidos del análisis. A partir de ese compromiso, en el Juzgado Decano de lo Penal se elaboró un documento general que debía ser completado por cada uno de los juzgados que decidiese colaborar en el estudio. Tal documento, que había de ser firmado por el juez, el secretario judicial —que es, en puridad, el encargado de la custodia del material derivado de los procedimientos—, así como por mí misma, contenía los compromisos siguientes:

Compromisos relativos a los datos personales

- a. *Esta investigación revelará la fuente de los datos (esto es, los juzgados penales de Barcelona que suministren documentación, tanto escrita como en soporte audiovisual).*
- b. *La seguridad, el anonimato y la privacidad de las personas que aparecen, directa o indirectamente, en los documentos que forman parte del corpus serán respetados rigurosamente.*
- c. *Se tendrá especial cuidado en la preservación de la confidencialidad de los datos personales y se guardará secreto sobre los datos de carácter personal a los que tenga acceso la investigadora.*
- d. *Se cuidará al máximo el derecho a la propia imagen de las personas que aparezcan en las grabaciones, de modo que el uso de estas será rigurosamente limitado al desarrollo de la investigación.*
- e. *Las grabaciones se transcribirán omitiendo cualquier dato personal que permita identificar a las personas que, directa o indirectamente, están implicadas.*
- f. *En ningún caso se harán copias de los documentos (escritos o audiovisuales) que forman el corpus, ni se permitirá el acceso a ellos a terceras personas.*

Compromisos relativos a la guardia y custodia, y a la destrucción del corpus

- a. *La investigadora pondrá especial cuidado en la guardia y custodia de las grabaciones y los documentos que formen parte del corpus durante todo el tiempo que dure la investigación.*
- b. *Una vez terminada la investigación, la investigadora destruirá todos los documentos que le hayan sido proporcionados por los juzgados penales antes citados.*

Compromisos relativos a la publicación de la investigación

En cualquier publicación en la que se emplee el corpus solicitado:

- a. *Se mencionará la ayuda para la obtención de tal corpus recibida por los juzgados penales de Barcelona;*
- b. *Se pondrá particular cuidado en la omisión de datos de carácter personal a fin de que la información contenida en el corpus quede disociada de cualquier persona identificada o identificable;*
- c. *Se incluirá la referencia explícita a la omisión de datos personales a fin de cumplir con el compromiso ético de este documento.*

Entre todos los procedimientos de los que obtuve documentación, seleccioné diez, de diferentes dimensiones y temas distintos, con la idea de confeccionar un corpus que fuera variado y abarcable.

4.2. Segunda etapa: la digitalización y compilación de los textos

Tras decidir cuáles de los documentos obtenidos iban a formar parte del CPP, se inició la etapa correspondiente a la digitalización de los textos. De entrada, se planteaba la cuestión de si era conveniente digitalizar todos los textos de forma completa o bien seleccionar, dado el propósito de la investigación, solamente los fragmentos narrativos de cada texto. Finalmente, opté por digitalizar e incluir en el corpus los documentos enteros de los diez procedimientos escogidos, con la excepción de los membretes institucionales del encabezamiento, así como la fecha y las firmas. Las razones para hacerlo son diversas. En primer lugar, es conveniente conservar el documento global para poder determinar qué lugar ocupan los textos narrativos en su seno y por qué motivo. En segundo lugar, resulta difícil decidir de antemano qué fragmentos tienen una dimensión narrativa y cuáles no, de modo que, eliminando aquellos que, a priori, pueda parecer que no intervienen en la configuración del relato, se corre el riesgo de obviar segmentos textuales que, tras el análisis, resulten importantes. Además, de cara a configurar un corpus de procesos penales que pueda ser utilizado en investigaciones futuras, parece más adecuado disponer de un conjunto de textos completos. Por otra parte, uno de los objetivos derivados de la composición del CPP es el de poder utilizarlo para la docencia en

facultades de derecho, de forma que, como es evidente, resulta más enriquecedor disponer de los textos por entero.

Respecto a la digitalización de los textos escritos que conforman el CPP (esto es, los escritos de las partes y las sentencias), el volcado a un formato digital no planteaba grandes dificultades. Se mantuvieron los errores de normativa y los enunciados agramaticales. En cambio, se prescindieron de los elementos tipográficos (negritas, cursivas, mayúsculas, etc.) de los documentos escritos originales. Se han conservado únicamente las letras mayúsculas al inicio de oración y en la primera letra de los nombres propios²⁵. En cuanto a la transcripción de los textos orales, se emplearon las convenciones de la tabla 1:

TABLA 1. CONVENCIONES DE TRANSCRIPCIÓN EMPLEADAS EN EL CPP

Aspectos prosódicos	
..	pausa (< 1 segundo)
...	pausa larga (> 1 segundo)
/	fin del grupo tonal
-	grupo tonal truncado
¿ ?	interrogativa convencional
?	inicio declarativo, final interrogativo
palabra	pronunciación enfática
<i>palabra</i>	pronunciación débil
aa	alargamiento vocálico
pp	alargamiento consonántico
< >	solapamiento
&	interrupción
" "	cita de palabras ajenas
Aspectos vocales: sonidos paralingüísticos	
mhm, aha	asentimiento
m::, e::, a::	vacilación antes de o durante un enunciado
uf:	duda
ntx	desacuerdo (clic alveolar)
Aspectos gestuales	
(afirmación)	gesto de afirmación con la cabeza
(negación)	gesto de negación con la cabeza
Otras	
xxx	fragmento ininteligible
[(?) palabra]	fragmento incierto
<i>palabra</i>	discurso en otro idioma
{comentario}	comentario de la transcriptor

25 Se han considerado nombres propios las referencias a los juzgados determinados (por ejemplo, Juzgado Penal nº 8 de Barcelona), así como las referencias a los textos normativos completos (como el Código Penal). En cambio, siguiendo la *Ortografía de la lengua española* (2010), en §4.6., sobre la mayúscula de relevancia, se ha decidido utilizar la minúscula para “ministerio fiscal” y “fiscal” –que, en contra de los usos habituales de los escritos jurídicos, se recogen ya en minúscula en la 22ª edición del *Diccionario de la lengua española*–, así como para “(su) señoría”.

Estos criterios están inspirados en propuestas de transcripción ortográfica de corpus orales clásicos del análisis de la conversación (Atkinson y Heritage 1984), así como en propuestas de confección de corpus orales del español y el catalán (Briz y grupo Val.Es.Co 2002, Payrató 1995), que se han simplificado, atendiendo a dos propósitos en conflicto. De un lado, se pretendía confeccionar un corpus adecuado y riguroso para la investigación científica. De otro lado, se ha perseguido con verdadero interés obtener un corpus que resulte máximamente legible, para que pueda ser explotado posteriormente en la capacitación comunicativa de juristas y futuros juristas.

Al margen de eso, como se ha apuntado en relación con el compromiso ético, una de las cuestiones éticas que debe plantearse el lingüista que emplea textos reales es el tratamiento de las referencias personales que aparecen en su corpus. Según plantea Rock (2001), con frecuencia, el investigador se debate entre la voluntad de mantener fielmente los datos que analiza y la voluntad de que los participantes/informantes no puedan ser identificados. Esta es una cuestión que la lingüística ha de considerar seriamente, en especial, cuando un corpus contiene elementos tan personales y se gestan en situaciones tan sensibles como las de un proceso judicial.

Por ello, el CPP ha sido anonimizado, esto es, el nombre y el resto de datos personales de todas las personas que aparecen en ellos han sido cambiados u ocultados (Rock 2001: 1, Thomson et al. 2005). Puesto que en ningún caso se solicitó a las personas implicadas su consentimiento para emplear la documentación del proceso en el cual participaron, es imperativo guardar el máximo rigor en la ocultación de nombres, datos personales, números de teléfono y otros detalles que permitan la identificación.

Como explica Rock (2001: 10), los criterios de anonimización deben establecerse para cada corpus concreto. Así, se ha descartado, por ejemplo, la opción de emplear únicamente las siglas de las personas implicadas, ya que esta opción, si bien es adecuada para otro tipo de estudios, dificulta la lectura de textos narrativos. El sistema de anonimización que se ha seguido en este trabajo²⁶ es este: se han substituido todos los nombres y apellidos por otros equivalentes (atendiendo a (i) si son de hombre o de mujer, (ii) su métrica y fonética parecidas,

26 Para una visión general sobre diferentes modos de anonimizar un corpus, véase Hasund (1994).

y (iii) si son socialmente equivalentes, esto es, se han buscado nombres comúnmente asociados a un mismo grupo social, de edad, región, etc.). El mismo criterio se ha seguido para la anonimización de otros nombres propios, como la denominación de las calles. Algo más compleja ha resultado la anonimización de las ciudades: en general, se ha mantenido el nombre de Barcelona; y se han substituido los nombres de otros municipios, por municipios cercanos al original y de características socio-económicas similares. Por último, en relación a las marcas comerciales y otros nombres propios (como el nombre de hospitales, por ejemplo), que, aunque en una medida mucho menor, también pueden contribuir a identificar a las personas concretas, se han sustituido por nombres métrica y fonéticamente similares; no obstante, se han conservado los nombres de marcas tan conocidas como Google o Sogecable, ya que emplear términos no conocidos podría hacer perder algunos matices de la historia y, desde luego, su uso frecuente hace que no sea factible individualizar una historia.

Finalmente, la compilación de los textos dio lugar a un corpus formado por diez procesos penales. Cada uno de ellos integra los géneros siguientes: (i) el escrito de la acusación pública, es decir, del ministerio fiscal; (ii) el escrito de la defensa –en los procesos en los que hay dos acusados (#5 y #10) hay dos escritos de defensa–; (iii) el juicio oral y (iv) la sentencia judicial; asimismo, en algunos procesos, además de la acusación pública, se persona la acusación particular. La composición del CPP se resume en la tabla siguiente:

TABLA 2: COMPOSICIÓN DEL CPP (TIPOS DE GÉNEROS POR PROCESOS)

Proceso	Escrito del ministerio fiscal	Escrito de acusación particular	Escrito de la defensa	Juicio oral	Sentencia
#1.	x		x	x	x
#2.	x		x	x	x
#3.	x	x	x	x	x
#4.	x		x	x	x
#5.	x		x (2)	x	x
#6.	x	x	x	x	x
#7.	x		x	x	x
#8.	x		x	x	x
#9.	x	x	x	x	x
#10.	x		x (2)	x	x

5 EL CORPUS DE PROCESOS PENALES

El resultado de la compilación es un corpus especializado compuesto por diez procesos penales que fueron resueltos por juzgados penales de la ciudad de Barcelona entre febrero de 2009 y julio de 2010. El cómputo total de palabras que conforman el CPP es de 98.943; la tabla 3 contiene el número de palabras concreto de cada uno de los géneros analizados, así como un cómputo total en cada proceso. Por su parte, la tabla 4 recoge la duración temporal de cada juicio oral y el cómputo global de la duración de todos los juicios del CPP (5h 59' 19").

TABLA 3: CÓMPUTO DE PALABRAS DEL CPP (POR GÉNERO Y POR PROCESO)

	Escrito del ministerio fiscal	Escrito de acusación particular	Escrito de la defensa	Juicio oral	Sentencia	Total palabras / proceso
#1.	787		335	4.968	1.936	8.026
#2.	571		263	2.044	2.072	4.950
#3.	621	673	261	3.271	5.525	10.351
#4.	513		395	3.967	1.425	6.300
#5.	542		425 + 393	2.949	1.014	4.930
#6.	911	748	265	1.484	949	4.357
#7.	543		218	1.602	1.086	3.449
#8.	675		543	10.869	2.221	14.308
#9.	418	3.227	401	15.494	3.188	22.728
#10.	1.023		1.569 + 1.107	11.470	4.375	19.544
Total palabras / corpus						98.943

TABLA 4: DURACIÓN DE CADA JUICIO ORAL Y CÓMPUTO GLOBAL DEL CPP

Duración del juicio oral	
#1.	30' 31"
#2.	11' 08"
#3.	19' 19"
#4.	23' 24"
#5.	20' 23"
#6.	11' 33"
#7.	8' 50"
#8.	1h 12' 27"
#9.	1h 33' 44"
#10.	1h 8'
Total	5h 59' 19"

La selección de estos procesos para la composición del CPP estuvo determinada por sus dimensiones (dado que se pretendía recoger procesos de tamaños diferentes) y por su temática. La tabla 5 recoge el motivo por el que se sigue cada uno de los procesos compilados:

TABLA 5: CAUSAS POR LAS QUE SE SIGUEN LOS PROCESOS DEL CPP

#1.	Delito de robo con violencia + falta de lesiones
#2.	Delito de quebrantamiento de condena
#3.	Delito de lesiones
#4.	Delito de simulación de delito
#5.	Delito de robo con fuerza en las cosas
#6.	Delito de amenazas en el ámbito familiar + falta de injurias
#7.	Delito contra la seguridad vial
#8.	Delito de tenencia ilícita de armas
#9.	Delito contra la propiedad intelectual
#10.	Delito de robo con violencia

Por otra parte, cabe dar cuenta de si los profesionales concretos que intervienen en los procesos del corpus son hombres o mujeres, pues algunos trabajos sobre comunicación en los juicios han puesto de manifiesto diferencias de género en el comportamiento comunicativo de los juristas (Tieszen y Pantoga 2006). Debido a que uno de los propósitos perseguidos con la elaboración del CPP es ponerlo a disposición de otros investigadores, podría resultar de interés para futuros trabajos conocer el género de los participantes en cada uno de los procesos. Con ese propósito, en la tabla siguiente se ha indicado qué operadores jurídicos de los procesos analizados son hombres (H) y cuáles son mujeres (M):

TABLA 6: CORRELACIÓN OPERADOR JURÍDICO-GÉNERO (HOMBRE-MUJER) EN EL CPP

	Juez	Fiscal	Abogado de la acusación particular	Abogado defensor	Perito	Agentes de policía
#1.	M	H	-	H	-	-
#2.	H	M	-	H	-	H, H
#3.	H	M	H	M	-	-
#4.	H	H	-	M	H, M	-
#5.	H	H	-	M	-	H, H, H, M
#6.	H	H	H	M	-	-
#7.	M	M	-	M	-	-
#8.	M	M	-	M	H, H	H, H
#9.	H	M	H	M	M, H	H, H, H
#10.	M	M	-	M	-	-

6 CONSIDERACIONES FINALES

Como se ha venido señalando, el estudio del discurso jurídico en lengua española adolece de falta de corpus suficientes, al alcance de cualquier investigador

interesado en la materia. Sería esperable que la creciente investigación en este ámbito produjese corpus cada vez mayores, más completos y abiertos a la comunidad científica. Ahora bien, la elaboración de corpus de español del derecho habría de estar acompañada por una reflexión metodológica adecuada. Así, resulta fundamental plantear cómo han de diseñarse y confeccionarse corpus apropiados para el análisis específico del español en contextos jurídicos.

La experiencia en la confección del CPP ha permitido identificar algunos aspectos que deben tenerse en cuenta a la hora de elaborar un corpus judicial. Para empezar, es imprescindible considerar las peculiaridades del sistema judicial en el que se centra la investigación y cómo condicionan el discurso; en ese sentido, el diseño de un corpus judicial debería atender a cuestiones como el sistema jurídico, el orden jurisdiccional, la instancia procesal, el tipo de procedimiento, etc., para generar un cuerpo de datos verdaderamente significativo e idóneo para el estudio del objeto analítico concreto que se desee abordar. Además, una de las características discursivas más destacadas de la práctica del derecho tiene que ver con el hecho de que en ella suelen implicarse multitud de hablantes, en posiciones diversas, que generan productos discursivos variados, emitidos de modos y en momentos distintos. Por ello, el estudio de la lengua jurídica –incluso cuando aborda específicamente un género discursivo en particular– no puede perder de vista que los elementos discursivos del derecho no son autónomos, sino que mantienen una relación esencial con otros elementos aparentemente distantes. En consecuencia, en el análisis del discurso del derecho es primordial atender a la amplia cadena de eventos comunicativos en la que se produce un fenómeno, y valorar en qué sentido el corpus para el estudio de ese fenómeno debe integrar tal cadena. Por último, no todos los corpus son aptos para cualquier tipo de análisis. En realidad, el tipo de análisis que se pretende llevar a cabo condiciona tanto la forma y el tamaño del corpus como el tratamiento que se le dé y la manera en que se explote.

Aparte de las necesidades empíricas de cada estudio, el diseño de un corpus debe tener en cuenta las dificultades reales a la hora de acceder a los textos. En ese sentido, resulta esperable también que la aparición creciente de estudios sobre discurso jurídico basados en corpus favorezca la cooperación entre los agentes jurídicos y las universidades. Es preciso, para ello, que se normalice la colaboración entre investigadores y profesionales del derecho. Debido a que,

con frecuencia, manejar textos jurídicos tiene implicaciones éticas, es conveniente promover la reflexión seria sobre cómo tratar los datos personales, la voz, las imágenes, etc. de los participantes en los eventos comunicativos que la Lingüística pretende analizar.

Este artículo ha pretendido abordar, de forma inicial, cuestiones que la investigación sobre discurso jurídico no debería seguir postergando. Las cuestiones y las propuestas que aquí se han presentado deberían ser completadas y mejoradas a partir de nuevas experiencias en la confección de corpus jurídicos. Solo de ese modo el estudio del español del derecho podrá constituirse en un área de estudio sólida, provista de una metodología propia y adecuada a sus fines.

REFERENCIAS BIBLIOGRÁFICAS

- ALCARAZ, Enrique y HUGHES, Brian (2002): *El español jurídico*. Barcelona: Ariel.
- ÁLVAREZ, Miriam (1995): *Tipos de escrito III: epistolar, administrativo y jurídico*. Madrid: Arco/Libros.
- ANDRÉS IBÁÑEZ, Perfecto (2003): “Sobre el valor de la intermediación. (Una aproximación crítica)”, *Jueces para la Democracia* 46: 57-66.
- ATKINSON, Maxwell y HERITAGE, John (eds.) (1984): *Structures of social action: studies in Conversational Analysis*. Cambridge: Cambridge UP.
- BAKER, Paul (2006): *Using corpora in discourse analysis*. Londres/Nueva York: Continuum.
- BAZERMAN, Charles (1994): “Systems of genres and the enactment of social intentions”, en FREEDMAN, Aviva y MEDWAY, Peter (eds.) *Genre and the new rhetoric*. Londres/Bristol: Taylor & Francis, 79-101.
- BHATIA, Vijay; LANGTON, Nicola y LUNG, Jane (2004): “Legal discourse: opportunities and threats for corpus linguistics”, en CONNOR, Ulla y UPTON, Thomas (eds.) *Discourse in the professions. Perspectives from corpus linguistics*. Ámsterdam/Philadelphia: John Benjamins, 203-231.
- BIBER, Douglas; CONNOR, Ulla y UPTON, Thomas (2007): *Discourse on the Move. Using corpus analysis to describe discourse structure*. Ámsterdam/Philadelphia: John Benjamins.

- BIBER, Douglas; CONRAD, Susan y REPPEN, Randi (1998): *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- BORDONABA, Maria Cristina (2009): “El lenguaje jurídico”, en CALVI, Maria Vittoria; BORDONABA, Cristina; MAPELLI, Giovanna y SANTOS, Javier: *Las lenguas de especialidad en español*. Roma: Carocci, 147-170.
- BOWEN, Glenn (2008): “Naturalistic inquiry and the saturation concept: a research note”, *Qualitative Research* 8(1): 137–152.
- BOWKER, Lynne y PEARSON, Jennifer (2002): Working with specialized language. *A practical guide to using corpora*. Londres/Nueva York: Routledge.
- BRIZ, Antonio y grupo Val.Es.Co (2002): *Corpus de conversaciones coloquiales*. Anejo de la revista Oralía. Madrid: Arco/Libros.
- CARRANZA, Isolda (2003): “Genre and institution: narrative temporality in final arguments”, *Narrative Inquiry* 13(1): 41-69.
- CARRANZA, Isolda (2008): “Metapragmatics in a courtroom genre”, *Pragmatics* 18(2): 169-188.
- CARRANZA, Isolda (2010): “La escenificación del conocimiento oficial”, *Discurso & Sociedad* 4(1): 1-29.
- CASANOVAS, Pompeu (1997): “Pragmática de las decisiones judiciales: diferencias de razonamiento y lenguaje ante el jurado”, *Cuadernos de Derecho Judicial* 16: 109-142.
- COTTERILL, Janet (ed.) (2002): *Language in the legal process*. Basingstone: Palgrave.
- COTTERILL, Janet (2003): *Language and power in court. A linguistic analysis of the O.J. Simpson trial*. Basingstone: Palgrave.
- CUBELLS, Jenny (2002): *La construcción social del delito: un estudio etnográfico en la práctica del Derecho Penal*. Tesis Doctoral. Universidad Autónoma de Barcelona.
- CUCATTO, Mariana (2009): “La conexión en las sentencias penales de primera instancia. Del análisis de textos a la práctica de escritura de sentencias”, *Revista de Llengua i Dret* 51: 135-160.
- DREW, Paul (1985): “Analyzing the use of language in courtroom interaction”, en VAN DIJK, Teun (ed.) *Handbook of discourse analysis (volumen 3: Discourse and dialogue)*. Londres: Academic Press, 133-147.

- DREW, Paul (1992): "Contested evidence in courtroom cross-examination: the case of a trial for rape", en DREW, Paul y HERITAGE, John (eds.) *Talk at work: interaction in institutional settings*. Cambridge: Cambridge UP, 470-520.
- FELTON ROSULEK, Laura (2008): "Manipulative silence and social representation in the closing arguments of a child abuse case", *Text & Talk* 28(4): 529-550.
- FELTON ROSULEK, Laura (2010): "Prosecution and defense closing speeches. The creation of contrastive closing arguments", en COULTHARD, Malcolm y JOHNSON, Alison (eds.): *The Routledge handbook of forensic linguistics*. Londres y Nueva York: Routledge, 218-230.
- FERREIRO, Xulio (2003): "Del procedimiento abreviado. Disposiciones generales", en PÉREZ-CRUZ, Agustín (coord.) *Las reformas del procedimiento abreviado, juicios rápidos y de faltas*. Granada: Comares, 1-53.
- FLOWERDEW, Lynne (2008): *Corpus-based analyses of the problem-solution pattern*. Ámsterdam: John Benjamins.
- FOUCAULT, Michel (1973): *Moi, Pierre Rivière, ayant égorgé ma mère, ma sœur et mon frère: un cas de parricide au XIXe siècle*. París: Gallimard.
- GARBERÍ, José (2000): "Título III: Del procedimiento abreviado para determinados delitos", en GIMENO SENDRA, Vicente; CONDE-PUMPIDO, Cándido y GARBERÍ, José, *Los procesos penales* (vol. 6). Barcelona: Bosch.
- GEWIRTZ, Paul (1996): *Narrative and rhetoric in the law*, en BROOKS, Peter y GEWIRTZ, Paul (eds.): *Law's stories. Narrative and rhetoric in the law*. New Haven/Londres: Yale UP, 2-13.
- GÓMEZ COLOMER, Juan Luis (2001): "El jurado español: ley y práctica", *Revue internationale de droit pénal* 72: 285-312.
- GÓMEZ MARTÍNEZ, Carlos (2003): "La grabación del sonido y de la imagen en los juicios civiles: del juez lector al juez espectador", *Jueces para la democracia* 48: 81-87.
- HARRIS, Sandra (2001): "Fragmented narratives and multiple tellers: witness and defendant accounts in trials", *Discourse studies* 3(1): 53-74.
- HARRIS, Sandra (2005): "Telling stories and giving evidence: the hybridisation of narrative and non-narrative modes of discourse in a sexual assault trial", en THORNBORROW, Joanna y COATES, Jennifer (eds.) *The sociolinguistics of narrative*. Ámsterdam: John Benjamins, 215-237.

- HARVEY, Penelope (1992): "Bilingualism in the Peruvian Andes", en CAMERON, Deborah; FRAZER, Elizabeth; HARVEY, Penelope; RAMPTON, M.B.H. y RICHARDSON, Kay: *Researching language. Issues of power and method*. Londres y Nueva York: Routledge, 65-89.
- HASUND, Kristine (1994): "Protecting the innocent: the issue of informants' anonymity in the COLT corpus", en RENOUF, Antoinette (ed.) *Explorations in corpus linguistics*. Ámsterdam/Atlanda: Rodopi, 13-28.
- HEFFER, Chris (2005): *The language of jury trial: a corpus-aided analysis of legal-lay discourse*. Basingstoke/Nueva York: Palgrave.
- HEFFER, Chris (2010): "Narrative in trial. Constructing crime stories in court", en COULTHARD, Malcolm y JOHNSON, Alison (eds.): *The Routledge handbook of forensic linguistics*. Londres/Nueva York: Routledge, 199-217.
- HILTUNEN, Risto (1996): "Tell me, be you a witch?: Questions in the Salem witchcraft trials of 1692", *International Journal for the Semiotics of Law* 9(1): 17-37.
- HILTUNEN, Risto (2004): "Salem 1692: a case of courtroom discourse in a historical perspective", en HILTUNEN, Risto y WATANABE, Shinichiro (eds.) *Approaches to style and discourse in English*. Osaka: Osaka UP, 3-21.
- HUNSTON, Susan (2002): *Corpora in Applied Linguistics*. Cambridge: Cambridge UP.
- HUNSTON, Susan (2008): "Collection strategies and design decisions", en LÜDELING, Anke y KYTÖ, Merja (eds.) *Corpus Linguistics: an international handbook* (vol.1.). Berlín/Nueva York: Walter de Gruyter, 154-168.
- IGARTUA, Juan (2003): *La motivación de las sentencias, imperativo constitucional*. Madrid: Centro de Estudios Políticos y Constitucionales.
- JACQUEMET, Marco (1996): *Credibility in court. Communicative practices in the camorra trials*. Cambridge: Cambridge UP.
- KENNEDY, Graeme (1998): *An introduction to Corpus Linguistics*. Londres/ Nueva York: Longman.
- KOESTER, Almut (2010): "Building small specialised corpora", en O'KEEFFE, Anne y MCCARTHY, Michael (eds.) *The Routledge Handbook of Corpus Linguistics*. Londres/Nueva York: Routledge, 66-79.
- LATORRE, Virgilio (2002): *Función jurisdiccional y juicios paralelos*. Madrid: Civitas.

- LEE, David (2008): "Corpora and discourse analysis. New ways of doing old things", en BHATIA, Vijay; FLOWERDEW, John y JONES, Rodney (eds.) *Advances in discourse studies*. Nueva York: Routledge: 86-99.
- LÓPEZ SAMANIEGO, Anna (2006a): "El uso metaargumentativo de las perífrasis obligativas en el lenguaje judicial español", en CALVI, Maria Vittoria y CHIERICHETTI, Luisa (eds.) *Nuevas tendencias en el discurso de especialidad*. Berna: Peter Lang, 117-140.
- LÓPEZ SAMANIEGO, Anna (2006b): "Los ordenadores del discurso enumerativos en la sentencia judicial: ¿estrategia u obstáculo?", *Revista de Llingua i Dret* 45: 61-87.
- LÓPEZ SAMANIEGO, Anna (2010): "Documentos profesionales con destinatarios no expertos. El empleo de los mecanismos referenciales en la sentencia del 11M", *Signos* 43(72): 99-123.
- LÜDELING, Anke y KYTÖ, Merja (2008): "Introduction", en LÜDELING, Anke y KYTÖ, Merja (eds.) *Corpus Linguistics: an international handbook* (vol.1.). Berlín/Nueva York: Walter de Gruyter, v-xii.
- MAGRO, Vicente (2010): "La grabación de los juicios penales en la reforma de la LECrim. por Ley 13/2009, de 3 de noviembre: la actuación del Secretario Judicial en los juicios penales y el principio de inmediación en materia de recursos", *La Ley Penal* 67: 6.
- MALEY, Yon (1994): "The language of the law", en GIBBONS, John (ed.) *Language and the law*. Londres/Nueva York: Longman, 11-50.
- MCENERY, Tony y WILSON, Andrew (1996): *Corpus Linguistics*. Edimburgo: Edinburg University Press.
- MCENERY, Tony; XIAO, Richard y TONO, Yukio (2006): *Corpus-based language studies: an advanced resource book*. Nueva York: Routledge.
- MILLER, Binny (2000): "Telling stories about cases and clients: the ethics of narrative", *Georgetown Journal of Legal Ethics* 14(1): 1-54.
- MONTERO AROCA, Juan (2000): "Sentencia, motivación y requisitos internos. La congruencia", en MONTERO AROCA, Juan; GÓMEZ COLOMER, Juan Luis; MONTÓN, Alberto y BARONA, Silvia (eds.) *El nuevo proceso civil* (Ley 1/2000). Valencia: Tirant lo Blanch, §20.
- MORENO CATENA, Víctor; COQUILLAT, Ángela; DE DIEGO, Luis Alfredo; JUANES, Ángel y DE LLERA, Emilio (2000): *El proceso penal*:

- doctrina, jurisprudencia y formularios* (vol. 4). Valencia: Tirant lo Blanch.
- MUERZA, Julio (2002): *El proceso penal abreviado*. Cizur Menor: Aranzadi.
- O'KEEFFE, Anne y MCCARTHY, Michael (eds.) (2010): *The Routledge Handbook of Corpus Linguistics*. Londres/Nueva York: Routledge.
- PASCUAL, Esther (2003): *Imaginary dialogues. Conceptual blending and fictive interaction in criminal courts*. Tesis doctoral, Universidad de Ámsterdam.
- PAYRATÓ, Lluís: "Transcripción del discurso coloquial", en CORTÉS, Luis (ed.) *El español coloquial*. Almería: Universidad de Almería, 45-70.
- PHILIPS, Susan (1987): "On the use of WH Questions in American courtroom discourse: a study of the relation between language form and language function", en KEDAR, Leah (ed.) *Power through discourse*. Norwood: Ablex, 83-113.
- POBLET, Marta (1998): *Las formas retóricas del discurso jurídico. Una descripción etnográfica*. Barcelona: Institut de Ciències Polítiques i Socials.
- REPPEN, Randi (2010): "Building a corpus. What are the key considerations?", en O'KEEFFE, Anne y MCCARTHY, Michael (eds.) *The Routledge Handbook of Corpus Linguistics*. Londres/Nueva York: Routledge, 31-38.
- RIDAO, Susana (2007): *Análisis pragmalingüístico de resoluciones de conflictos: las mediaciones laborales. Propuestas de investigación*. (Tesis doctoral). Almería: Universidad de Almería.
- RIDAO, Susana (2009a): "El interrogatorio en los contextos judiciales: análisis del tipo y la frecuencia de las preguntas utilizadas en los juzgados de lo penal de Almería capital", *Interlingüística* 18: 940-950.
- RIDAO, Susana (2009b): *Estrategias de (des)cortesía en las mediaciones laborales*. Murcia: Universidad de Murcia.
- ROCK, Frances (2001): "Policy and practice in the anonymisation of linguistic data", *International Journal of Corpus Linguistics* 6(1): 1-26.
- SAMANIEGO, Eva (2004): "El lenguaje jurídico: peculiaridades del español jurídico", en FUERTES, Pedro (coord.) *Lengua y sociedad: investigaciones recientes en Lingüística Aplicada*. Valladolid: Universidad de Valladolid.
- SINCLAIR, John (2005): "Corpus and text. Basic principles", en WYNNE, Martin (ed.) *Developing linguistic corpora*. Oxford: The Oxford Text Archive: 1-21.
- STOREY-WHITE, K. (1997): "Kissing the jury: the advantages and limitations

- of the “keep it simple” principle in the presentation of expert evidence of courts and juries”, *Forensic Linguistics* 4(2): 280-287.
- STUBBS, Michael (1996): *Text and corpus analysis: computer-assisted studies of language and institutions*. Oxford: Blackwell.
- STYGALL, Gail (1994): *Trial language. Differential discourse processing and discourse formation*. Ámsterdam/Philadelphia: John Benjamins.
- SVARTVIK, Jan (ed.) (1992): *Directions in Corpus Linguistics: proceedings of Nobel symposium*. Berlín: Mouton de Gruyter.
- TARANILLA, Raquel (2006): “Consideraciones sobre la polifonía en el escrito de declaración del atestado policial”, en NEUMANN, Claus-Peter; PLO, Ramón y PÉREZ-LLANTADA, Carmen (coords.) *Actas del V Congreso Internacional AELFE*. Zaragoza: Publicaciones de la Universidad de Zaragoza, 70-76.
- TARANILLA, Raquel (2007): “*Con cuentos a la policía*: las secuencias narrativas en el acta de declaración del atestado policial”, *Revista de Llengua i Dret* 47: 79-112.
- TARANILLA, Raquel (2009): “La gestión de la propia imagen en las argumentaciones del Tribunal Constitucional: la función retórica de las estrategias de cortesía”, *Revista de Llengua i Dret* 52: 117-149.
- TARANILLA, Raquel (2011a): “Análisis lingüístico de la transcripción del relato de los hechos en el interrogatorio policial”, *Estudios de Lingüística*. Universidad de Alicante 25.
- TARANILLA, Raquel (2011b): *La configuración narrativa en el proceso penal. Un análisis discursivo basado en corpus*. Tesis doctoral: Universitat de Barcelona.
- TARANILLA, Raquel (2012): *La justicia narrante. Un estudio sobre el discurso de los hechos en el proceso penal*. Cizur Menor: Aranzadi.
- TARUFFO, Michele (2008): “¿Verdad negociada?”, *Revista de Derecho* XXI(1): 129-152.
- THOMSON, Denise; BZDEL, Lana; GOLDEN-BIDDLE, Karen; REAY, Trish y ESTABROOKS, Carole (2005): “Central questions of anonymization: a case study of secondary use of qualitative data”, *Forum: Qualitative Social Research* 6(1): art. 29 [consulta on-line: 20.12.2010, <http://www.qualitative-research.net/index.php/fqs/article/viewArticle/511/1102>]
- TIERSMA, Peter (2006): *Communicating with juries: how to draft more unders-*

- tandable jury instructions*. Williamsburg: National Center for State Courts.
- TIESZEN, Bozena y PANTOGA, Heather (2006): “Gender-based miscommunication in legal discourse and its impact on the clarity of legal language”, *International Journal for the Semiotics of Law* 19: 69-80.
- TOGNINI-BONELLI, Elena (2001): *Corpus Linguistics at work*. Ámsterdam: John Benjamins.
- TORRUELLA, Joan y LLISTERRI, Joaquim (1999): “Diseño de corpus textuales y orales”, en BLECUA, José Manuel; CLAVERÍA, Gloria; SÁNCHEZ, Carlos y TORRUELLA, Joan (eds.) *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Milenio, 45-77.
- VIDMAR, Neil y HANS, Valerie (2007): *American juries: the verdict*. Amherst: Prometheus.
- VIRTANEN, Tuija (2008): “Corpora and discourse analysis”, en LÜDELING, Anke y KYTÖ, Merja (eds.) *Corpus Linguistics: an international handbook* (vol.2.). Berlín/Nueva York: Walter de Gruyter, 1043-1070.
- WALTER, Bettyruth (1988): *The jury summation as speech genre*. Ámsterdam/ Filadelfia: John Benjamins.

Fecha de recepción: 14 de mayo de 2012

Fecha de aceptación: 7 de marzo de 2013

