

EL IMPACTO DEL DERECHO EN EL USO DE LAS TECNOLOGÍAS DEL LENGUAJE EN LAS ADMINISTRACIONES PÚBLICAS*

Agustí Cerrillo i Martínez**

Raquel Xalabarder***

Resumen

Las administraciones públicas han incorporado las tecnologías del lenguaje como instrumento para prestar los servicios públicos de manera más eficaz y eficiente. Asimismo, las administraciones públicas están fomentando, a través de diversas medidas, la industria de las tecnologías del lenguaje. Este artículo tiene por objetivo identificar las normas que deben ser tenidas en cuenta al impulsar el diseño, desarrollo y uso de las infraestructuras lingüísticas en las administraciones públicas. Además, también persigue valorar el impacto de la normativa para impulsar el uso de las tecnologías del lenguaje en las administraciones públicas.

Palabras clave: Tecnologías del lenguaje; derecho; propiedad intelectual; protección de datos; recursos lingüísticos; administraciones públicas.

THE IMPACT OF LAW ON THE USE OF LANGUAGE TECHNOLOGIES IN PUBLIC ADMINISTRATIONS

Abstract

Public administrations have incorporated language technologies as an instrument to provide more effective and efficient public services. The language technologies industry is also being promoted by public administrations through various measures. This article aims to identify the rules that should be taken into account when promoting the design, development and use of language infrastructures in public administrations and seeks to assess the impact of such regulations.

Keywords: Language technologies; law; intellectual property; linguistic resources; public administrations.

* Este trabajo resulta del proyecto Análisis normativo para el uso y la apertura de recursos lingüísticos en el marco del Plan de Impulso de las Tecnologías del Lenguaje elaborado por los autores por encargo de Red.es.

** Agustí Cerrillo, catedrático de derecho administrativo en la Universitat Oberta de Catalunya. acerrillo@uoc.edu

*** Raquel Xalabarder, catedrática de propiedad intelectual en la Universitat Oberta de Catalunya. rxalabarder@uoc.edu

Artículo recibido el 24.05.2018. Evaluación ciega: 16.10.2018 y 25.10.2018. Fecha de aceptación de la versión final: 03.12.2018.

Citación recomendada: Cerrillo, Agustí; Xalabarder, Raquel (2018). El impacto del derecho en el uso de las tecnologías del lenguaje en las administraciones públicas. *Revista de Llengua i Dret, Journal of Language and Law*, (70), 17-30, DOI: [10.2436/rid.i70.2018.3159](https://doi.org/10.2436/rid.i70.2018.3159).

Sumario

- 1 Introducción
- 2 El rol de la Administración pública en el fomento de las tecnologías del lenguaje
 - 2.1 Los datos abiertos de interés lingüístico
 - 2.2 La reutilización de recursos lingüísticos
 - 2.3 La interoperabilidad de los recursos y aplicaciones lingüísticas
 - 2.4 La normativa lingüística
 - 2.5 El derecho de la competencia
- 3 La propiedad intelectual
 - 3.1 La gestión de la propiedad intelectual en los recursos de interés lingüístico
 - 3.2 La protección de los procesadores y plataformas de interés lingüístico
- 4 La protección de los datos personales y la anonimización
- 5 La confidencialidad de la información
- 6 Reflexiones finales

1 Introducción

Las tecnologías del lenguaje son aquellas tecnologías utilizadas para la comprensión automática del lenguaje humano. Generalmente, bajo esta expresión se incluyen las tecnologías de procesamiento de lenguaje natural y las tecnologías de traducción automática, que pueden ser utilizadas por las administraciones públicas para la traducción automática de documentos, la realización de resúmenes de resoluciones, el análisis de informes o la respuesta automática a preguntas formuladas por la ciudadanía en lenguaje natural en los servicios de atención.

Las tecnologías del lenguaje se basan en el uso de **recursos lingüísticos** consistentes en textos, audios o vídeos cuyo contenido se sujeta a tratamiento lingüístico. Cualquier procesamiento del lenguaje natural se basa en el uso de grandes volúmenes de recursos lingüísticos múltiples y variados, de muy distinto origen y tipología, tales como *corpus* de textos (textos científicos, documentos traducidos y memorias de traducción, contratos, sentencias, historias clínicas, patentes, textos anotados, etc.), listas de nombres (de personas, organizaciones, marcas, topónimos, etc.), taxonomías y clasificaciones o terminologías y diccionarios.¹ Estos recursos pueden ser **de titularidad pública o privada** y **provenir de fuentes distintas**.

Los recursos son objeto de “**tratamiento**” diverso, tales como análisis, indexación e interrelación, dando lugar a corpus de recursos “metadateados” y glosarios o memorias de traducción. Son utilizados a través de **procesadores** (generalmente, programas de ordenador y bases de datos, que pueden ser desarrollados por la propia Administración pública o de titularidad privada) para realizar distintas operaciones y obtener resultados diversos, como la traducción automática o la minería de datos (*text and data mining*). Tanto los recursos como los procesadores se hallan normalmente protegidos por regímenes de propiedad intelectual e industrial o quedan sujetos a acuerdos de confidencialidad.² A su vez, para evaluar la calidad de los recursos lingüísticos y los procesadores, se llevan a cabo campañas de evaluación que permiten conocer el grado de calidad y de precisión de los resultados obtenidos. La última etapa del procesamiento del lenguaje natural conlleva la reutilización propia —por ejemplo, para realizar campañas de evaluación— o ajena de tales infraestructuras, para enriquecer los recursos con nuevos tratamientos y procesamientos, dando lugar a nuevos servicios o nuevas informaciones basados en ellos.

Los recursos lingüísticos, los procesadores y las campañas de evaluación conforman lo que se conoce como **infraestructuras lingüísticas**.

El uso de las infraestructuras lingüísticas por parte de las administraciones públicas genera **múltiples cuestiones jurídicas**. Este trabajo tiene por objetivo identificar las normas que deben ser tenidas en cuenta al impulsar el diseño, desarrollo y uso de las infraestructuras lingüísticas en las administraciones públicas.

2 El rol de la Administración pública en el fomento de las tecnologías del lenguaje

Las administraciones públicas pueden jugar un rol esencial en el impulso de los proyectos de tecnologías del lenguaje facilitando la reutilización de los recursos o de los procesadores lingüísticos que están bajo su poder. La puesta a disposición de las infraestructuras lingüísticas para su reutilización debe realizarse en unos formatos técnicos y bajo unas condiciones jurídicas que faciliten su utilización y garanticen la **interoperabilidad** entre recursos y procesadores, a fin y efecto de poder extraer el máximo provecho posible del procesamiento del lenguaje.

2.1 Los datos abiertos de interés lingüístico

Los recursos lingüísticos que generan las administraciones públicas pueden difundirse en formatos que faciliten su reutilización.³ En particular, los datos abiertos son todos aquellos datos en poder de las administraciones

1 Asimismo, los recursos no solo pueden estar en formato de texto sino que eventualmente también se pueden utilizar como recursos vídeos, imágenes o fotografías.

2 Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (2015: 21).

3 En este marco, el Plan de Impulso de las Tecnologías de la Información prevé que en el marco de la política de reutilización de la información del sector público deben introducir el concepto de datos abiertos de interés lingüístico y difundir este concepto en la Administración pública.

públicas que se caracterizan desde un punto de vista técnico por ser procesables electrónicamente y, desde el punto de vista jurídico, por la ausencia de restricciones que limiten su reutilización (propiedad intelectual, datos personales, etc.).

La modificación operada por la Ley 18/2015, de 9 de julio, de la Ley 37/2007, de 16 de noviembre, de reutilización de la información del sector público (en adelante, LRISP), fomenta la difusión de datos en formatos abiertos al disponer que “las administraciones y organismos del sector público facilitarán sus documentos en cualquier formato o lengua preexistente, pero también procurarán, siempre que ello sea posible y apropiado, proporcionarlos en formato abierto y legible por máquina conforme a lo previsto en el apartado anterior y conjuntamente con sus metadatos, con los niveles más elevados de precisión y desagregación. Tanto el formato como los metadatos, en la medida de lo posible, deben cumplir estándares y normas formales abiertas”.⁴

De todos modos, las administraciones públicas no tienen la obligación de difundir los datos en dicho formato ni tampoco están obligados a crear documentos, adaptarlos o facilitar extractos de documentos, cuando ello suponga un esfuerzo desproporcionado que conlleve algo más que una simple manipulación.⁵

En cualquier caso, las administraciones públicas deben difundir la información en aplicación de lo previsto en la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (en adelante, LTAIBG), en formatos reutilizables.⁶ En particular, por lo que respecta al portal de transparencia de la Administración General del Estado, se fomentará que la información sea publicada en formatos que permitan su reutilización.⁷

2.2 La reutilización de recursos lingüísticos

La normativa sobre reutilización de la información del sector público fomenta y regula el uso de informaciones, datos y documentos que están en poder de las administraciones y organismos del sector público, por parte de personas físicas o jurídicas, con finalidades comerciales o no comerciales, siempre que este uso no constituya una actividad administrativa pública ni suponga el intercambio de documentos entre administraciones públicas en el ejercicio de las funciones públicas que tengan atribuidas.⁸

La LRISP recoge la obligación para las administraciones y organismos del sector público de autorizar la reutilización de documentos (en relación con lo previsto en el considerando 8 de la Directiva 2013/37/UE).⁹ Sin embargo, una lectura atenta del artículo 4 permite observar ciertas limitaciones en el alcance de dicha obligación. En efecto, en dicho artículo se dispone que “los documentos de las administraciones públicas y organismos del sector público serán reutilizables en los términos que prevé esta Ley”, y se remite a los regímenes establecidos en el mismo artículo, que son determinados por cada Administración pública. Así pues, el alcance de esta obligación de facilitar la reutilización viene determinado en el mismo precepto que prevé que “las administraciones públicas y organismos del sector público velarán porque los documentos a los que se aplica esta normativa puedan ser reutilizados para fines comerciales o no comerciales”.¹⁰ Por tanto, será la regulación de la reutilización de cada Administración pública la que en última instancia determinará la posible reutilización como recursos lingüísticos de los datos que están en su poder.

4 Artículo 5.2 LRISP. La modificación de la LRISP en esta materia responde a lo previsto en el artículo 5 de la Directiva 2003/98/CE, relativa a la reutilización de la información del sector público —modificada en este punto por la Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013.

5 Artículo 5.2 LRISP.

6 Artículo 5.4 LTAIBG.

7 Artículo 11.c LTAIBG.

Las normas autonómicas de transparencia han sido más explícitas en relación al impulso de los datos abiertos, lo que puede tener un impacto positivo a la hora de promover los datos abiertos lingüísticos previendo la difusión de los datos públicos en formatos abiertos, la creación de portales de datos abiertos o la creación de catálogos de información reutilizable.

8 Artículo 3 Ley 37/2007, de 16 de noviembre, de reutilización de la información del sector público. La LRISP es de aplicación en las administraciones y organismos del sector público. Sin embargo, el hecho de que una entidad no esté incluida en el ámbito subjetivo de la LRISP no es obstáculo para que pueda facilitar la reutilización de la información que produce o está en su poder.

9 Respecto al régimen jurídico de la reutilización, véase Valero Torrijos (2012).

10 Artículo 4.2 LRISP.

La LRISP define como documento toda información, sea cual sea su soporte y forma de expresión, elaborada o custodiada por las administraciones públicas y organismos públicos.¹¹ De este modo, todos aquellos datos que hayan producido o estén en poder de las administraciones u organismos públicos son susceptibles de convertirse en recursos lingüísticos al servicio de los desarrolladores y usuarios de las tecnologías del lenguaje. Sin embargo, la LRISP excluye del régimen de reutilización previsto en ella algunos documentos, como aquellos que no sean accesibles, aquellos cuyo conocimiento requiera una legitimación especial y los contenidos públicos creados, por ejemplo, por entidades de radiodifusión o entidades educativas.¹²

Tal y como se ha indicado en el apartado anterior, para facilitar la reutilización, las administraciones públicas deben proporcionar los datos en formato abierto y legible por máquina por medios electrónicos y mediante plataforma multicanal y conjuntamente con sus metadatos, con los niveles más elevados de precisión y desagregación y, cuando sea posible, cumpliendo estándares y normas formales abiertas.¹³ Asimismo, las administraciones públicas deben difundir los datos para su reutilización a través de espacios específicamente dedicados a ello como portales de datos abiertos donde se puede acceder a *corpus* u otros recursos lingüísticos.¹⁴ Junto a ello, también puede ser de utilidad poner a disposición de desarrolladores y usuarios de las tecnologías del lenguaje catálogos que integren los datos o recursos públicos reutilizables.¹⁵

En el caso de que se decida facilitar la reutilización, la LRISP define tres modalidades de reutilización: reutilización sin sujeción a condiciones; reutilización con sujeción a condiciones, que pueden venir establecidas en licencias tipo; y reutilización con solicitud previa, que puede incorporar condiciones en una licencia.

Con frecuencia, la reutilización de los recursos lingüísticos quedará sujeta al cumplimiento de determinadas condiciones, que deben ser claras, justas y transparentes; no pueden restringir las posibilidades de reutilización ni limitar la competencia y además no pueden ser discriminatorias para categorías comparables de reutilización.¹⁶ En esta dirección, la LRISP recoge algunas condiciones que se pueden establecer al autorizar la reutilización de información del sector público, tales como que el contenido de la información, incluyendo sus metadatos, no sea alterado; no se desnaturalice el sentido de la información; se cite la fuente; se mencione la fecha de la última actualización; la finalidad o finalidades concretas para las que es posible la reutilización futura de los datos cuando la información contenga datos de carácter personal o la prohibición de revertir el procedimiento de disociación.¹⁷

Además, en tanto los recursos lingüísticos sujetos a reutilización estén protegidos por la propiedad intelectual, también deberán establecerse licencias de propiedad intelectual según lo que se indicará posteriormente.

2.3 La interoperabilidad de los recursos y aplicaciones lingüísticas

Los recursos lingüísticos pueden tener orígenes diversos y pueden encontrarse en formatos diferentes. Para facilitar su reutilización es recomendable que los formatos que se utilicen sean interoperables. Para garantizar la interoperabilidad de los recursos lingüísticos es necesario establecer unas normas, criterios y recomendaciones que permitan a las aplicaciones utilizadas por las administraciones públicas, así como por los usuarios de los procesadores del lenguaje, intercambiar información y utilizar mutuamente la información intercambiada.

La interoperabilidad de los recursos lingüísticos es un aspecto de gran relevancia para facilitar el desarrollo del procesamiento del lenguaje natural.¹⁸ En la actualidad, la Norma Técnica de Interoperabilidad de reutilización de recursos de la información establece el conjunto de pautas básicas para la reutilización de

¹¹ Anexo LRISP.

¹² Artículo 3.3 LRISP.

¹³ Artículo 5 LRISP.

¹⁴ Así lo dispone específicamente para la Administración General del Estado el artículo 1.2 del Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la LRISP.

¹⁵ En esta dirección, el artículo 4.5 LRISP.

¹⁶ Artículo 4.3 LRISP.

¹⁷ Artículo 8 LRISP.

¹⁸ Así, lo ha puesto de manifiesto la Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, que ha modificado en este extremo la Directiva de 2003 (artículo 5).

documentos y recursos de información elaborados o custodiados por el sector público y las condiciones comunes sobre selección, identificación, descripción, formato, condiciones de uso y puesta a disposición de los documentos y recursos de información elaborados o custodiados por el sector público.

La misma exigencia de interoperabilidad es predicable de los sistemas y aplicaciones de procesamiento del lenguaje natural. La Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público (en adelante, LRJSP), persigue promover la reutilización de sistemas y aplicaciones de propiedad de la Administración y dispone que las administraciones públicas deben poner a disposición de cualquier Administración pública que lo solicite las aplicaciones que bien hayan desarrollado sus propios servicios técnicos, bien hayan sido objeto de contratación y de cuyos derechos de propiedad intelectual sean titulares.¹⁹ Para facilitar la reutilización y la interoperabilidad, la LRJSP también prevé que las aplicaciones pueden ser declaradas de fuentes abiertas.²⁰

2.4 La normativa lingüística

Al hablar de la regulación del proceso del lenguaje natural no puede desconocerse el papel que en este ámbito debe jugar la normativa lingüística en aquellas comunidades autónomas que disponen de lengua propia.

Sin embargo, únicamente la Ley 1/1998, de 7 de enero, de política lingüística de Cataluña, prevé explícitamente que el gobierno de la Generalidad debe favorecer, estimular y fomentar la investigación, producción y comercialización de todo tipo de productos en catalán relacionados con las industrias de la lengua, como son los sistemas de reconocimiento de voz, de traducción automática y similares u otros posibles de acuerdo con los avances tecnológicos.

2.5 El derecho de la competencia

Tanto el fomento del desarrollo de la industria de las tecnologías del lenguaje natural como la oferta de servicios vinculados a dichas tecnologías por parte de las propias administraciones públicas, puede generar problemas desde la perspectiva del derecho de la competencia. Al respecto, cabe tener presente lo previsto en la normativa de defensa de la competencia que exige garantizar que las ayudas o subvenciones que se faciliten a la industria para el desarrollo de las tecnologías del lenguaje no puedan ser contrarias a las normas de la competencia ni a nivel estatal ni tampoco comunitario.

Asimismo, en la oferta en el mercado de las aplicaciones desarrolladas por las administraciones públicas, es necesario diferenciar esta actividad de aquella estrictamente relacionada con el ejercicio de sus funciones públicas. Cuando las administraciones públicas decidan ofrecer servicios en el mercado basados en las aplicaciones desarrolladas por ellas debe garantizarse la libre competencia respecto a desarrolladores de aplicaciones, evitando de este modo el abuso de posición dominante, los acuerdos exclusivos o las ayudas cruzadas u otras prácticas que puedan ser consideradas contrarias a la competencia. En cualquier caso, la dificultad en este punto radica en delimitar adecuadamente estas funciones.

3 La propiedad intelectual

La normativa de propiedad intelectual se aplica tanto a contenidos de titularidad pública como privada. Con frecuencia, los **recursos y procesadores** sujetos a reutilización (a los efectos de tratamiento lingüístico) por parte de la Administración pueden estar protegidos por la propiedad intelectual.

De hecho, el régimen de la propiedad intelectual afecta a los proyectos de tecnologías del lenguaje en todas sus diferentes etapas: tanto en el uso inicial de recursos preexistentes para su tratamiento y en el desarrollo de aplicaciones para llevar a cabo tal tratamiento (programa de ordenador), como en la reutilización de los resultados que se deriven de tales tratamientos y, en concreto, en el establecimiento de las condiciones (licencias) de reutilización.

¹⁹ Artículo 157 LRJSP. Todo esto está previsto y desarrollado en el Esquema Nacional de Interoperabilidad y en la Norma Técnica de Interoperabilidad.

²⁰ Artículo 157.2 LRJSP.

Antes de iniciar cualquier proyecto de procesamiento lingüístico es necesario identificar si el material a utilizar está sujeto al régimen de protección de la propiedad intelectual y, de ser así, si existe la debida autorización (ya sea legal o contractual) para realizar tal tratamiento.

Para saber qué contenidos están protegidos por el régimen de propiedad intelectual (y cuales no), y qué actos deben quedar sujetos a la previa autorización del titular (y cuales no), debemos remitirnos a tal normativa.

3.1 La gestión de la propiedad intelectual en los recursos de interés lingüístico

El régimen de la propiedad intelectual confiere al **autor** un conjunto de **derechos** sobre su **obra**, por un **plazo de protección** determinado y sujeto a algunas **excepciones o límites** legalmente previstos.

Los conceptos de autor y de obra son, pues, los pilares fundamentales e indisolubles del régimen de propiedad intelectual: no hay obra sin autor, ni autor sin obra. Obra es toda creación original *expresada*. Basta un *quantum* mínimo de originalidad (no todo está protegido), pero incluso entonces sólo se protege la expresión: el derecho de autor no protege las ideas, ni la información, ni los datos. Tampoco se protegen la normativa, las resoluciones judiciales y los actos de organismos públicos, ni las traducciones oficiales de todos ellos. Además, puesto que el derecho de autor sólo se protege por un plazo determinado,²¹ podrán ser libremente reutilizables aquellos recursos que se hallen en el **dominio público** por haber expirado ya su plazo de protección.

Por ello, en tanto se trate de creaciones originales (obras) u otras prestaciones²² todavía protegidas, el uso de recursos para el tratamiento lingüístico deberá respetar la normativa de propiedad intelectual. En cambio, cuando el tratamiento lingüístico lo sea de obras o prestaciones en el dominio público, o de datos o información no protegida, o de normativa, resoluciones judiciales o actos oficiales de organismos públicos excluidos de protección, el tratamiento podrá hacerse libremente, sin necesidad de contar con la previa autorización de su titular.

Ahora bien, a pesar de que los datos y la información puedan ser utilizables libremente, en toda la Unión Europea se reconoce al **fabricante de una base de datos** el derecho a autorizar o prohibir tanto la copia (extracción) como la explotación (reutilización) de la totalidad o una parte sustancial de tales contenidos, y ello a pesar de no estar estos protegidos por el derecho de autor. Este derecho “sui generis” puede ser especialmente relevante para los proyectos de tratamiento del lenguaje natural: por ejemplo, cuando se utilizan recursos extraídos de una base de jurisprudencia o normativa o de una base de datos “puros” (esto es, datos meteorológicos, datos sanitarios o expedientes administrativos). A pesar de que tales contenidos no estén protegidos por la propiedad intelectual, el fabricante de la base de datos que los contiene podrá autorizar o prohibir su extracción y reutilización.

En todo caso, el tratamiento lingüístico de recursos (protegidos o no) puede quedar sujeto a **condiciones concretas de reutilización** establecidas en las fuentes de donde se obtienen: por ejemplo, a través de una base de datos (ya sea de suscripción o de acceso abierto) o de una página web de datos abiertos de la Administración pública. Téngase en cuenta que cuando se trata de contenidos de acceso abierto, las condiciones de explotación (o reutilización) pueden venir detalladas como “condiciones de uso” o “aviso legal” en el propio sitio web.

Cuando el tratamiento y procesamiento lo sea de recursos protegidos por el régimen de propiedad intelectual, se realizarán diversos “**actos de explotación**” que exigen la previa autorización de su autor o titular. En todo caso, se estarán realizando actos de reproducción (hacer copias) y de comunicación pública (por ejemplo, difundir por internet); también pueden realizarse actos de distribución (si el tratamiento y la explotación de los recursos lingüísticos se realiza mediante ejemplares tangibles, en papel, CD o DVD). Además, cuando el tratamiento y procesamiento de recursos protegidos conlleve la creación de otros recursos que constituyan,

21 La regla general, en toda la Unión Europea, es de 70 años *post mortem auctoris* (PMA), pero en España debe tenerse en cuenta un plazo más largo (80 años PMA) aplicable a obras de autores fallecidos antes del 7 diciembre 1987.

22 El texto refundido de la Ley de Propiedad Intelectual protege también a los artistas (actores, cantantes, músicos, etc.), a los productores (musicales y audiovisuales), a los realizadores de “meras” fotografías (que no llegan a alcanzar la categoría de “obra” fotográfica) y a los editores de ciertas producciones editoriales.

a su vez, objeto de propiedad intelectual, necesitaremos obtener también autorización para realizar actos de “transformación” de tales recursos.

Ahora bien, antes de que la obra o prestación entre en el dominio público, el texto refundido de la Ley de Propiedad Intelectual (TRLPI) recoge un listado de **excepciones o límites** que permiten ciertos actos de explotación (por ejemplo, el llamado “derecho de cita”, la parodia y la copia privada). Estos límites son la manifestación de la función social que, como toda propiedad, tiene la propiedad intelectual. Identificar claramente el alcance de estos límites legales es fundamental para saber cuándo el tratamiento o procesamiento lingüístico de un recurso protegido queda directamente autorizado por ley y cuando exigirá la preceptiva autorización o licencia.

Lamentablemente, ninguno de los límites legalmente previstos ofrece suficiente amparo legal para poder realizar actos de procesamiento del lenguaje, prescindiendo de la correspondiente licencia del titular. El límite de copia privada, por ejemplo, permitiría a los ciudadanos (no a las empresas ni a la Administración pública) realizar copias de obras, pero no transformarlas (realizar el tratamiento) ni reutilizarlas (ponerlas a disposición del público); de forma similar, el límite a favor de bibliotecas, hemerotecas o archivos, sólo para fines “de investigación o de conservación” dejaría fuera a muchos proyectos de procesamiento lingüístico.²³ Tampoco el límite de las llamadas “copias técnicas” podría dar cobertura a todos los actos de explotación del procesamiento lingüístico.²⁴ Estrechamente relacionado con el tratamiento lingüístico, existe una propuesta de límite específico a favor de la minería de datos que actualmente está siendo considerada por la Comisión y el Parlamento de la Unión Europea.²⁵ De aprobarse con la redacción de la propuesta inicial, este límite autorizaría la reproducción y extracción de recursos protegidos para fines de minería de datos, siempre y cuando se trate de contenidos a los que se tenga acceso legal y la minería se realice para fines de investigación científica por parte de organismos de investigación. Es decir, se trata de un límite muy reducido que difícilmente permitiría entender autorizados gran parte de los actos de procesamiento lingüístico analizados: ya que el procesamiento lingüístico realizado para fines que no sean propiamente de investigación científica quedarían *de facto* sujetos al ámbito de exclusividad del autor.

Hasta aquí hemos examinado como afecta el régimen de propiedad intelectual a la utilización de recursos protegidos para realizar acciones de tratamiento y procesamiento lingüísticos. Este mismo régimen será de aplicación a los resultados de dichas acciones. Cuando el tratamiento y procesamiento lingüístico dé lugar a resultados suficientemente originales para ser considerados, a su vez, obras protegidas por el derecho de autor (por ejemplo, algunas memorias de traducción), estaremos ante una nueva obra, protegida por la propiedad intelectual y cuya autoría y titularidad se registrará por las normas generales: la titularidad de los derechos de propiedad intelectual sobre los *corpus* resultantes del tratamiento lingüístico será del autor o coautores que lo han realizado.

El autor es la persona natural que crea la obra²⁶ y es titular de los derechos que le permiten controlar “en exclusiva” la explotación de su obra: la facultad absoluta de autorizar y prohibir actos de explotación. La Administración podrá ser titular de estos derechos exclusivos cuando los obtenga por cesión expresa (por

²³ Artículos 31.2 y 37.1 TRLPI, respectivamente.

²⁴ Artículo 31.1 TRLPI. Es posible que alguna de las acciones de tratamiento lingüístico cumpliera las exigentes condiciones allí previstas, quedando autorizadas por ley; pero es previsible que muchas de las copias que se realizan para el procesamiento lingüístico no las cumplirán y requerirán, por tanto, la correspondiente autorización. Además, este límite solo abarcaría los actos de reproducción, quedando los subsiguientes actos de comunicación pública o de transformación de los recursos utilizados sujetos a autorización.

²⁵ Véase el artículo 3 de la Propuesta de Directiva sobre los derechos de autor en el Mercado Único Digital COM(2016) 593 final.

²⁶ Artículo 5 TRLPI. Para ser autor hay que crear, no basta con hacer (la pericia técnica —por compleja y valiosa que sea— no es suficiente para otorgar la condición de autor). Cuando en la creación de una obra participan varias personas, hablaremos de coautoría y de obra en colaboración (artículo 7 TRLPI) o colectiva (artículo 8 TRLPI) —la distinción entre ambas dependerá del grado de colaboración entre los autores y de la existencia de una iniciación, coordinación y divulgación externa.

ejemplo, en las cláusulas de la licitación y/o en el contrato de encargo)²⁷ o mediante la presunción de cesión legalmente prevista para la obra creada por funcionario o trabajador asalariado.²⁸

La LRISP dispone que el ejercicio de los derechos de propiedad intelectual de las administraciones y organismos del sector público deberá realizarse “de forma que se facilite su reutilización”. En concreto, el Real Decreto de desarrollo de la LRISP establece, para la Administración General del Estado, el uso de cesiones gratuitas y no exclusivas de los derechos de propiedad intelectual necesarios para desarrollar la actividad de reutilización autorizada, en cualquier modalidad y bajo cualquier formato, para todo el mundo y por el plazo máximo permitido por la Ley. Así pues, en la medida en que los recursos y resultados del procesamiento lingüístico sean objeto de protección por la propiedad intelectual, la Administración deberá establecer una licencia de propiedad intelectual para autorizar su reutilización. Antes, es fundamental asegurar que la Administración pública está legitimada para sujetar tales contenidos a la reutilización y asegurar la “cadena de título”.

Cuando la Administración es titular de derechos de propiedad intelectual (por ejemplo, los recursos generados por la propia Administración pública), podrá decidir libremente si sujetar tales contenidos a la reutilización y establecer las condiciones de tal reutilización. Sin embargo, con frecuencia la Administración puede tener en su custodia contenidos (documentos) que quedan también sujetos a los mandatos de la LRISP (por ejemplo, contenidos custodiados en bibliotecas museos y archivos) pero sobre los que la Administración no ha adquirido ningún derecho de propiedad intelectual (por ejemplo, informes o fotografías aportados por los administrados como parte de un procedimiento administrativo). Para poner tales contenidos a disposición del público para su reutilización, la Administración deberá contar con la debida autorización del titular de derechos, mediante licencia o autorización.²⁹

La misma conclusión es predicable respecto de contenidos sobre los que recaigan derechos de propiedad industrial (patentes, marcas o diseños industriales) que sean titularidad de terceros —aunque no serán objeto de estudio en el presente trabajo. Si los recursos sujetos al tratamiento lingüístico estaban protegidos por el derecho de autor o conexos, la reutilización de los resultados requerirá la previa autorización de los titulares de los recursos preexistentes. Sin esta autorización previa, no será posible sujetar los resultados del tratamiento a la reutilización por parte del público.

Un formato de licencias de propiedad intelectual que ha obtenido gran aceptación pública, especialmente en internet y en el ámbito de la reutilización de la información del Sector Público (RISP), son las **licencias Creative Commons (CC)**, con las que el autor (o titular) autoriza al “público” en general a realizar actos de explotación de su obra, sujetos a unas condiciones que el autor escoja (esto es, finalidades comerciales y/o transformación). Cuando un usuario decide utilizar una obra bajo una licencia CC se convierte en licenciatarario y se compromete a aceptar y respetar las condiciones de la licencia establecida por el autor.

Las licencias CC pueden ser útiles no sólo para identificar aquellos recursos autorizados para tratamiento lingüístico, sino también para que la Administración autorice el tratamiento lingüístico de los recursos de su titularidad.

Las licencias Creative Commons pueden ser de gran utilidad, tanto por su difusión y fácil localización, como por la flexibilidad que ofrecen a la hora de ajustar el alcance de actos de explotación que se autorizan. Así pues, a la hora de determinar qué licencia de propiedad intelectual es la que mejor se ajusta al objetivo de la reutilización, la Administración pública deberá tener en cuenta que las más permisivas son las que mejor

²⁷ En tales casos, es fundamental que el contrato de encargo (o, previamente, en las cláusulas de la licitación) se incluya una cesión de derechos a favor de la Administración. Véase el artículo 308 de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público, así como las consideraciones realizadas ya en reutilización de activos. *Guía del Ministerio de Hacienda y Administraciones Públicas* (2015), apartados 46 a 52.

²⁸ Cuando la obra ha sido creada en virtud de una relación laboral o funcional, la ley prevé directamente una presunción de cesión, salvo pacto en contra, de los derechos de explotación a favor del empresario (o Administración) con el alcance necesario para poder realizar su actividad habitual. Artículo 51 TRLPI.

²⁹ Esta licencia (de derechos de explotación) puede ser de múltiples formas y obtenerse a través de canales diversos: ya sea directamente del titular de derechos o a través de las entidades de gestión colectiva, que autorizan la explotación de obras y prestaciones en nombre de los titulares (autores, artistas, productores) que les han encargado la gestión de sus derechos: por ejemplo, CEDRO (para las obras literarias), SGAE (para las obras musicales y audiovisuales) o VEGAP (para las obras de arte plástico). Artículo 147 y siguientes TRLPI.

facilitan la reutilización (BY), mientras que las que no permiten los fines comerciales (NC) o, especialmente, la adaptación y transformación (ND) son las que menos contribuyen a la reutilización. Otras opciones son las licencias CC0, pensadas justamente para licenciar la explotación de contenidos generados por la propia Administración y que ni siquiera exigen la atribución de autoría (como exige la licencia CC-BY) o incluso la posibilidad de establecer una licencia estándar propia de la Administración.

3.2 La protección de los procesadores y plataformas de interés lingüístico

Un programa de ordenador es un conjunto de instrucciones que hacen que una máquina —el ordenador— funcione y ejecute una tarea o proporcione un resultado. Los programas informáticos a través de los que se realiza el tratamiento del lenguaje tienen un régimen especial de protección dentro de la propia ley de propiedad intelectual.³⁰

Como vimos, autor es quien crea la obra (en este caso, el programa de ordenador); pero la ley prevé aquí una regla un tanto excepcional: el empresario que coordine la creación de un programa de ordenador en régimen de obra colectiva, será considerado autor del mismo y titular inicial de todos los derechos de propiedad intelectual.³¹

Buena parte de los procesadores de lenguaje natural que utilizan las administraciones públicas han sido desarrollados por encargo a la industria. Si la Administración que encarga el programa quiere ser titular de los derechos de explotación, deberá obtenerlos por cesión expresa (contrato) del autor / titular original; a ser posible, esta cesión de derechos debe incluirse ya en las cláusulas de la licitación, además de en el contrato de encargo. Desde un punto de vista general, en este punto debe hacerse una remisión a lo previsto en la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público (LCSP). En particular, la LCSP prevé que los contratos de servicios que tengan por objeto el desarrollo y la puesta a disposición de productos protegidos por un derecho de propiedad intelectual o industrial llevarán aparejada la cesión de estos derechos a la Administración pública contratante. Esta presunción de cesión será de aplicación en tanto en los pliegos de cláusulas administrativas o en el documento contractual no se establezca o acuerde otra cosa. En cualquier caso, se hayan cedido o no los derechos de propiedad intelectual o industrial a la Administración comitente, ésta podrá autorizar el uso del procesador a todos los entes, organismos y entidades que conforman el sector público.³²

A su vez, la explotación de un programa de ordenador puede hacerse a través de una gran variedad de licencias. La opción de las licencias de usuario final (*End User License Agreements*), que utiliza el llamado “software propietario” (p. ej., Microsoft), convierte a los usuarios en licenciatarios sin permitirles realizar ningún acto de explotación (salvo el de utilizar el programa de ordenador). En cambio, las licencias públicas, por ejemplo, la General Public License (GPL), permiten la libre reproducción, transformación y subsiguiente difusión del software, pero obligan al llamado “copyleft”: es decir, el programa resultante debe quedar sujeto a la misma licencia GPL, evitando así la “apropiación” del software y asegurando su subsiguiente reutilización.³³

Estas licencias deberán tenerse debidamente en cuenta, tanto al utilizar software preexistente para la creación del procesador o infraestructura lingüística, como para ponerlos a disposición del público para su reutilización.

También existen otras licencias públicas, como la Licencia Pública de la Unión Europea (EURL), que permiten la libre reutilización del software obligando simplemente a proporcionar el código fuente, pero sin obligar al “copyleft” (como hace la GPL); con lo cual permiten ajustar mejor los objetivos de la reutilización

30 El programa de ordenador incluye el código fuente y el código objeto, incluyendo la interfaz. Todo ello puede quedar protegido por el TRLPI, cuando sea una “creación intelectual propia de su autor”. También se protege la documentación preparatoria, la documentación técnica y los manuales de uso del programa de ordenador.

31 Artículo 97 TRLPI

32 Así lo dispone el artículo 157 LRJSP al prever que “las administraciones pondrán a disposición de cualquiera de ellas que lo solicite las aplicaciones, desarrolladas por sus servicios o que hayan sido objeto de contratación y de cuyos derechos de propiedad intelectual sean titulares”.

33 Es decir, si se utiliza código fuente licenciado con una GPL para realizar un nuevo procesador o infraestructura lingüística, ésta deberá sujetarse (en su totalidad) a la licencia GPL, y si se elige la GPL para autorizar la reutilización de procesadores o infraestructuras, los nuevos programas deberán también ser licenciados con la misma licencia GPL, negando así la posibilidad de explotarlos mediante sistemas de software “propietario” o licencias de usuario final (EULAs).

de la información del sector público con las expectativas de explotación de resultados subsiguientes por parte de quien invierte en su obtención.

4 La protección de los datos personales y la anonimización

Cuando los recursos lingüísticos objeto de tratamiento incluyan datos personales, deberán adoptarse las medidas legalmente previstas para asegurar un tratamiento adecuado de tales datos personales y respetar la privacidad de las personas afectadas.

Como punto de partida, debemos tener presente que se consideran datos personales cualquier información concerniente a una persona física identificada o identificable, es decir, cuya identidad pueda determinarse, directamente o indirectamente, en particular, a través de un número de identificación o uno o diversos elementos específicos, característicos de su identidad física, fisiológica, psíquica, económica, cultural o social. Algunos datos personales cuentan con una protección legal más elevada (por ejemplo, los datos relativos a la ideología, la afiliación sindical, la religión, el origen racial o étnico, la salud o la vida sexual).³⁴

Al procesar recursos lingüísticos que incluyan datos personales se estará realizando un tratamiento de datos personales, que quedará sometido a un conjunto de principios.³⁵ Este tratamiento está sujeto a diversas normas y principios de debido cumplimiento para asegurar la calidad de los datos (los datos personales deben ser adecuados, pertinentes y no excesivos respecto a la finalidad para la que se obtienen y también deben ser exactos y actualizados), el consentimiento del interesado, la seguridad (se deben adoptar las medidas de índole técnica y organizativa necesarias para garantizar la seguridad de los datos de carácter personal y evitar su alteración, pérdida, tratamiento o acceso no autorizado, así como garantizar la confidencialidad, integridad y disponibilidad de los datos); y el secreto en el tratamiento de los datos, así como la congruencia y racionalidad en su utilización.³⁶

Así pues, cuando se lleve a cabo el procesamiento de textos que contengan datos personales, deberán extremarse las precauciones y considerar debidamente los riesgos que pueda conllevar para sus titulares.³⁷ Cuando una Administración pública permite la reutilización de recursos que contengan datos de carácter personal, está realizando una cesión o comunicación de datos personales, es decir, una revelación de datos realizada a una persona distinta del interesado; para ello es necesario o bien contar con el consentimiento del interesado o bien que así lo prevea una ley (por ejemplo, cuando se trate de datos recogidos de fuentes accesibles al público, o cuando la cesión se produzca entre administraciones públicas y tenga por objeto el tratamiento posterior de los datos con fines históricos, estadísticos o científicos o se obtengan o elaboren los datos personales con destino a otra Administración).

En todo caso, la puesta a disposición del público para su reutilización de recursos que contengan datos personales, en poder de la Administración, deberá realizarse garantizando el pleno respeto de los principios que consagran la protección de datos personales. Por ejemplo, la reutilización se puede llevar a cabo si previamente se produce la disociación de los datos que impida la identificación de las personas. La disociación de los datos personales es aquel tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable, es decir, que permita obtener un dato disociado.³⁸ En este caso, estaríamos hablando de una cesión de datos anonimizados que no requiere el consentimiento

34 Artículo 9 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos (en adelante, RGPD).

35 Es decir, una operación o procedimiento técnico de carácter automatizado, que permite la recogida, grabación, conservación, elaboración, modificación, bloqueo y cancelación, así como la cesión de datos que resulte de comunicaciones, consultas, interconexiones y transferencias.

36 Véase al respecto Puyol Montero (2016).

37 Artículo 35 RGPD.

38 De acuerdo con la Agencia Española de Protección de Datos, hay que garantizar la irreversibilidad de la anonimización, y para ello deberán valorar tanto las fuentes de información disponibles en los diferentes medios y, en particular, a través de internet, como la tecnología aplicable en los procesos de anonimización y en los de reidentificación (Agencia Española de Protección de Datos, 2016). La anonimización se puede llevar a cabo utilizando diferentes técnicas. Véase al respecto Grupo de Trabajo sobre protección de datos del artículo 29 (2014).

de los interesados.³⁹ A pesar de la anonimización de los datos, las técnicas del *big data* permiten el cruce de datos de diferente origen, cosa que puede llevar a que finalmente se llegue a identificar a una persona, lo que exige ser especialmente cauteloso respecto a determinadas categorías de datos personales a disposición de las administraciones públicas como, por ejemplo, los datos sanitarios.⁴⁰

Así pues, cuando una Administración pública se plantee la posibilidad de facilitar la reutilización de recursos lingüísticos que contengan datos personales, debería llevar a cabo con carácter previo una evaluación del impacto relativa a la protección de datos.⁴¹ Por otra parte, en la medida en que el tratamiento de datos personales pudiera justificarse como “necesario para el cumplimiento de una misión realizada en interés público” (artículo 6.1.e RGPD) o “necesario para la satisfacción de intereses legítimos del responsable del tratamiento” (artículo 6.1.f RGPD), cabría —previa valoración del riesgo y el impacto que pudiera tener el tratamiento— entenderlo autorizado para ciertos objetivos concretos de reutilización (por ejemplo, para fines de tratamiento lingüístico), siempre salvaguardando —en la medida de lo posible— la anonimidad de las personas cuyos datos personales se sujetan al tratamiento/RISP.⁴²

Además de la anonimización técnica de los datos, se puede y debe garantizar la anonimización a través del establecimiento de condiciones en las licencias de reutilización, por ejemplo, cláusulas de protección de datos relativas a datos anonimizados o cláusulas relativas a las finalidades compatibles con la finalidad para la que los datos fueron inicialmente recabados.

5 La confidencialidad de la información

La confidencialidad de la información constituye una excepción al principio general de reutilización de los recursos públicos.⁴³

La LTAIBG ha incluido diferentes límites que son aplicables tanto a las obligaciones de difusión activa de la información como a las solicitudes de acceso a la información.⁴⁴ La concurrencia de uno de estos límites no supone la imposibilidad ni de acceder a la información ni de reutilizarla, siendo necesario que el acceso a la información (o su reutilización) pueda suponer un perjuicio para los bienes o derechos que dichos límites persiguen proteger. En todo caso, la aplicación de los límites debe ser justificada y proporcionada a su objeto y finalidad de protección y también deberá atender a las circunstancias del caso concreto, especialmente a la concurrencia de un interés público o privado superior que justifique el acceso.⁴⁵

En la actualidad, las administraciones públicas disponen de ingentes cantidades de información relativas a las empresas. Una parte de esta información se refiere a aspectos organizativos, de funcionamiento, sobre los productos o los servicios que presta la empresa o sobre los procesos que se siguen, cuyo conocimiento por terceros puede tener una incidencia negativa en el desarrollo de la actividad de la empresa y en su posición en el mercado. De este modo, es importante tener presentes los diversos límites previstos en la legislación para evitar que cualquier tratamiento de lenguaje natural pueda llegar a afectar a la competitividad de una empresa o a sus secretos empresariales.⁴⁶

39 Artículo 11.6 y, en particular en el ámbito de la salud, el artículo 16.3 de la Ley 41/2002.

40 Así lo advierte el Grupo de Trabajo sobre protección de datos del artículo 29 (2014). Asimismo, véase Martínez Martínez (2017).

41 Artículo 35 RGPD.

42 A los efectos de conseguir este equilibrio de intereses, véanse los criterios previstos en el artículo 6.4 RGPD para valorar la compatibilidad del tratamiento de datos para fines distintos del que fueron recogidos. Un análisis en general en Adsuara Varela (2016).

43 En esta dirección, la LRISP dispone que no será aplicable a los documentos sobre los que existan prohibiciones de acceso según lo previsto en la LTAIBG. Véase el resto de supuestos en el artículo 3.3 LRISP.

44 Artículo 14 LTAIBG.

45 Sobre los límites de la transparencia pública, véase Araguàs Galcerà (2017).

46 Para determinar el alcance de esta protección, téngase presente lo previsto en la Directiva (UE) 2016/943 del Parlamento Europeo y del Consejo, de 8 de junio de 2016, relativa a la protección de los conocimientos técnicos y la información empresarial no divulgados (secretos comerciales) contra su obtención, utilización y revelación ilícitas. Véase al respecto Girona Domingo (2018).

6 Reflexiones finales

A lo largo de las páginas anteriores se ha podido constatar cómo el derecho incide significativamente en el procesamiento del lenguaje. En efecto, más allá de lo que en la actualidad permiten la tecnología o la lingüística, las normas jurídicas determinan qué recursos lingüísticos se pueden utilizar o qué procesamientos se pueden realizar.

Tal y como se ha observado, la legislación vigente facilita en términos generales el uso de las tecnologías del lenguaje y atribuye una amplia flexibilidad para fomentar el uso y desarrollo de las tecnologías del lenguaje, decidiendo qué recursos lingüísticos se pueden utilizar o qué se puede hacer con los resultados obtenidos por el procesamiento de dichos recursos tanto por las propias administraciones públicas, como por parte de la industria de las tecnologías del lenguaje o la universidad.

Sin embargo, también se ha podido constatar que la regulación vigente puede dificultar, sino frenar, determinados procesos lingüísticos. En particular, se han apuntado algunas limitaciones que puede generar la legislación vigente, por ejemplo, en materia de propiedad intelectual (en relación a la regulación de la minería de datos y textos), y de protección de datos personales (por lo que respecta a la necesidad de recabar el consentimiento de los titulares).

Encontrándose ambas normas en pleno proceso de reforma en la Unión Europea y en España, respectivamente, se brinda una buena oportunidad para llamar la atención sobre la necesidad de contemplar específicamente las necesidades que las tecnologías del lenguaje requieren y ser consciente del impacto que el derecho puede tener en su legítimo desarrollo.

7 Referencias bibliográficas

Adsuara Varela, Borja (2016). El consentimiento. En José Luis Piñar Mañas (ed.), *Reglamento general de protección de datos: hacia un nuevo modelo europeo de privacidad* (pp. 151-170). Madrid: Reus.

Agencia Española de Protección de Datos (2016). *Orientaciones y garantías en los procedimientos de anonimización de datos personales*. Accesible en: <https://www.aepd.es/media/guias/guia-orientaciones-procedimientos-anonimizacion.pdf>

Araguàs Galcerà, Irene; Pérez García, Ignacio Luis; Artigas Monge, Raquel; Molina Roig, Isabel; Mancilla i Muntada, Francesc; Contreras Soler, Beatriz; Romera Santiago, Núria; Serrat Romani, Marina *et al.* (2017). *Los límites al derecho de acceso a la información pública*. Madrid: Instituto Nacional de Administración Pública.

Girona Domingo, Ramón Miguel (2018). Hacia una nueva protección de los secretos industriales y comerciales: La Directiva 2016/943 del Parlamento europeo y del Consejo. *Revista jurídica de la Comunidad Valenciana: jurisprudencia seleccionada de la Comunidad Valenciana*, (66), 5-20.

Grupo de Trabajo sobre protección de datos del artículo 29 (2014). *Dictamen 05/2014 sobre técnicas de anonimización*.

Martínez Martínez, Ricard (2017). Big data, investigación en salud y protección de datos personales: ¿Un falso debate? *Revista Valenciana d'Estudis Autònoms*, (62), 235-280.

Puyol Montero, Francisco Javier (2016). Los principios del derecho a la protección de datos. En José Luis Piñar Mañas (ed.), *Reglamento general de protección de datos: hacia un nuevo modelo europeo de privacidad* (pp. 135-150). Madrid: Reus.

Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (2015). *Plan de Impulso de las Tecnologías del Lenguaje*.

Valero Torrijos, Julián (2012). El acceso y la reutilización de la información administrativa: Implicaciones jurídicas del proceso de modernización tecnológica de las Administraciones Públicas en su actual y futura configuración. *Diario La Ley*, (7800).