



El uso de corpus electrónicos para la investigación de terminología jurídica

Los corpus lingüísticos digitales pueden ser recursos de gran utilidad para el estudio y la investigación de terminología en el ámbito de la traducción especializada, incluido el sector jurídico. En este artículo, exploramos los diversos corpus jurídicos disponibles para su consulta en la web, la posibilidad de compilación de otros propios y el uso de herramientas informáticas específicas para el análisis de corpus electrónicos de cualquier tipo en la labor de investigación lingüística.

| Por la **traductora pública Nadia Florencia Bernardi**, integrante de la Comisión de Recursos Tecnológicos

Los estudios lingüísticos siempre se han servido de la compilación de textos como herramienta de investigación de los diversos fenómenos de la lengua. Los corpus, como recopilaciones extensas de textos que constituyen muestras representativas de una lengua o una variedad lingüística en particular, son elementos indispensables en cualquier labor investigadora relacionada, ya que brindan datos lingüísticos reales que reflejan el uso de los términos en contexto.

Durante las últimas décadas, este enfoque metodológico tradicional tuvo un desarrollo acelerado gracias a los grandes avances tecnológicos producidos en el ámbito de la informática y, en especial, internet. La web se ha convertido en una inmensa fuente de la que se pueden extraer corpus de todo tipo gracias a su fácil acceso, y, en consecuencia, el proceso de compilación de textos se ha simplificado y agilizado. Tanto es así que, en la actualidad, el concepto de corpus se ha convertido prácticamente en sinónimo de corpus electrónico, ya que la mayoría de ellos se presentan ahora en formato digital y se analizan mediante aplicaciones informáticas específicas, lo que ha multiplicado de manera exponencial sus posibilidades de estudio y explotación.



La actividad del traductor siempre ha estado vinculada, de alguna manera u otra, a la recopilación y al análisis de documentos para su labor profesional y para la investigación de terminología, ya sea en forma de corpus propiamente dicho o como fuente de organización conceptual o terminológica. El corpus permite ir más allá del mero análisis descontextualizado de un término y ofrece la posibilidad de profundizar en la comprensión de las relaciones conceptuales, léxicas y pragmáticas que se establecen en los diversos tipos de discurso. Resulta particularmente útil en el campo de la traducción especializada, ya que contribuye a llenar numerosas lagunas terminológicas o conceptuales y hace posible resaltar los matices lingüísticos del discurso especializado, caracterizado por respetar una serie de convenciones lingüísticas, funcionales y textuales asociadas a la comunidad profesional.

En el ámbito de la traducción jurídica, donde los traductores (muchos sin ser abogados) deben producir textos que se ajusten a las características propias del discurso legal, los corpus electrónicos resultan de gran utilidad. De esta forma, la compilación de corpus específicos o el análisis de aquellos ya existentes constituyen, junto con la consulta en diccionarios y glosarios especializados, un instrumento de investigación terminológica casi imprescindible en todo trabajo de traducción legal. Por lo tanto, en este artículo, proponemos una serie de corpus jurídicos ya compilados que se encuentran disponibles en la web, así como algunas herramientas informáticas útiles para el análisis de cualquier corpus electrónico.

Corpus jurídicos en la web

En internet, podemos encontrar múltiples corpus jurídicos, ya sean monolingües, bilingües o multilingües, de diferentes instituciones y organizaciones, tanto nacionales como internacionales.

En el ámbito de la Argentina, el más completo, por su alcance nacional y su exhaustividad documental, es la recopilación de documentos legislativos, judiciales, doctrinales y administrativos del Sistema Argentino de Información Jurídica o SAIJ (<http://www.saij.gob.ar>), dependiente de la Secretaría de Justicia del Ministerio de Justicia y Derechos Humanos de la Nación. Dicha entidad también administra el sitio web InfoLeg (<http://www.infoleg.gob.ar>), un subcorpus legislativo conformado por leyes, decretos, resoluciones, decisiones administrativas y otras disposiciones que se publican en el Boletín Oficial de la República Argentina, y ofrece, asimismo, otros recursos lingüísticos, como el Banco de Vocabularios Jurídicos de Argentina (<http://vocabularios.saij.gob.ar>).

En el ámbito internacional, por un lado, uno de los corpus jurídicos multilingües elaborados con mayor exhaustividad es la compilación de documentos legislativos, judiciales y administrativos emitidos por la Comisión, el Consejo, el Parlamento y el Tribunal de Justicia de la Unión Europea, (EUR-Lex, <https://eur-lex.europa.eu/homepage.html>), entre otros, en sus veinticuatro lenguas oficiales. Un ejemplo similar, por su alcance internacional y su precisión lingüística, es el corpus paralelo de las Naciones Unidas (<https://conferences.unite.un.org/uncorpus>), un cuerpo de textos jurídicos, registros oficiales y otros documentos parlamentarios de dicha organización que son de dominio público y que se encuentran disponibles en sus seis lenguas oficiales.

Por otro lado, uno de los corpus jurídicos monolingües en inglés más completos es el corpus diacrónico del Old Bailey (<https://www.oldbaileyonline.org>) del Tribunal

El uso de corpus electrónicos para la investigación de terminología jurídica



Penal Central de Londres, que contiene documentos y registros judiciales de casi doscientos mil juicios llevados a cabo entre 1674 y 1913. En el ámbito estadounidense, encontramos también el Corpus of US Supreme Court Opinions (<https://www.english-corpora.org/scotus/>), que contiene miles de decisiones emitidas por la Corte Suprema de los Estados Unidos entre 1790 y la actualidad.

Por su parte, el sitio web del Boletín Oficial de Canadá (<http://www.gazette.gc.ca>) ofrece un corpus jurídico bilingüe donde se pueden consultar, tanto en inglés como en francés, todos los textos legislativos promulgados en dicho país desde 1998. Otro ejemplo de un corpus jurídico bilingüe en el ámbito internacional es el Bononia Legal Corpus o BoLC (http://corpora.dslo.unibo.it/bolc_eng.html), que aún se encuentra en su versión piloto. Si bien se concibió como un corpus jurídico multilingüe comparable, por el momento, solo incluye textos en inglés, representativos del sistema jurídico anglosajón, y textos en italiano, representativos del ordenamiento jurídico de base romanística.

Estos corpus mencionados constituyen herramientas de gran utilidad para la investigación terminológica y la labor del traductor, pero, tal como puede apreciarse, en general, se refieren a determinados géneros jurídicos, como el legislativo. Es decir, se dejan a un lado otros, como el contractual y el notarial, ya que los documentos pertenecientes a estas áreas son de naturaleza privada y, muchas veces, están protegidos por disposiciones de confidencialidad, lo cual dificulta la posibilidad de reunirlos en un corpus para su análisis.

Por lo tanto, en lo que respecta a la investigación terminológica de textos de este tipo, podemos consultar corpus de modelos de contratos o formularios legales, disponibles en múltiples sitios web que recopilan documentación jurídica, como <https://modelos-de-contratos.blogspot.com> o <http://www.freelegalforms.net>, o, incluso, podemos compilar nuestros propios corpus especializados a partir de documentos disponibles en la web. Otra solución es crear corpus bilingües con textos de estos géneros jurídicos a partir de nuestras propias traducciones mediante programas de traducción asistida por computadora, como SDL Trados, Wordfast o memoQ, ya que las memorias de traducción son, por definición, corpus paralelos, es decir, conjuntos de textos electrónicos en dos o más lenguas, que están alineados a nivel de párrafo, oración o frase.





Herramientas informáticas para el análisis de corpus

En general, las funciones de búsqueda disponibles en los programas de lectura de documentos digitales están diseñadas para buscar contenidos y no datos lingüísticos, por lo que presentan múltiples limitaciones a la hora de realizar una investigación terminológica. Estas deficiencias, sumadas a la gran cantidad de documentos que debemos analizar al trabajar con corpus, condujeron al desarrollo de herramientas informáticas específicas para la investigación lingüística, que permiten procesar una multiplicidad de documentos. Estos programas facilitan la búsqueda de concordancias lingüísticas y posibilitan el estudio exhaustivo de los distintos elementos del discurso.

Algunas de estas herramientas informáticas se pueden utilizar con corpus propios ya compilados, como WordSmith Tools (<https://www.lexically.net/wordsmith>) y AntConc (<http://www.laurenceanthony.net/software/antconc>); en tanto que otras aplicaciones se utilizan directamente en la web y no admiten ningún control sobre la selección de textos en los que se efectúa la búsqueda, como KWicFinder (<https://www.kwicfinder.com>) y WebCorp (<http://www.webcorp.org.uk/live>). Otros programas informáticos especializados, como LancsBox (<http://corpora.lancs.ac.uk/lancsbox>), ofrecen ambas posibilidades de búsqueda, es decir, funcionan tanto con corpus propios como con corpus previamente determinados.



En general, estas herramientas para el análisis de corpus tienen algunas funciones en común, con sus respectivas variaciones. En primer lugar, cuentan con una herramienta de concordancia, que hace posible conocer las concordancias lingüísticas de todas las instancias en las que la palabra objeto de búsqueda aparece en el corpus. Así, podemos ver cómo se utilizan habitualmente algunas palabras y frases en un conjunto determinado de textos. En segundo lugar, cuentan con una herramienta de listado de palabras, que muestra todas las que figuran en los textos en forma de lista e indica con cuánta frecuencia lo hace cada una y en cuántos archivos. Esto nos permite conocer rápidamente las palabras que tienen mayor frecuencia en un conjunto de textos determinado. Asimismo, cuentan con una herramienta de palabras clave, que muestra las que son inusualmente frecuentes en el corpus objeto de estudio, en comparación con otro corpus de referencia, y con cuánta frecuencia aparece cada una. Con esta función podemos identificar las palabras características del corpus como parte de un género textual, por ejemplo.

En conclusión, los corpus electrónicos ofrecen múltiples posibilidades de investigación a todos los profesionales que se dediquen a una actividad lingüística, sean traductores, terminólogos o revisores. Constituyen una herramienta práctica y ágil que facilita la investigación de terminología en el ejercicio de la traducción especializada, como en el ámbito jurídico, ya que ofrecen datos extraídos directamente de documentos especializados y permiten identificar y analizar estructuras, frases y términos en su contexto natural de producción, así como rasgos estilísticos específicos del sector. Son una herramienta más con la que cuenta el traductor jurídico para facilitar su labor, desempeñarse mejor y ofrecer a sus clientes un trabajo preciso y de gran calidad. ■