

---

## DERECHO, LINGÜÍSTICA COMPUTACIONAL Y NUEVAS TECNOLOGÍAS

---

**Emmanuela Sassi**

*Licenciada en lenguas modernas. Investigadora del Instituto di Linguistica Computazionale del CNR-Italia.*

Objetivo de esta comunicación es mostrar cómo en los últimos cincuenta años la tecnología informática ha conferido un notable impulso a la lingüística aplicada, modificando muchos de los presupuestos técnicos y metodológicos que antes constituían la base de los estudios lingüístico-literarios, y cómo esto contribuye a las aplicaciones posibles en la informática jurídica.

En consecuencia, los medios informáticos, que en sus comienzos parecían dirigidos al uso exclusivo de las disciplinas científicas, se han convertido en un instrumento cotidiano vistas sus aplicaciones que van del *word processing* o tratamiento de textos —que todo el mundo emplea para escribir hasta cartas personales—, al tratamiento automático del lenguaje y hasta la inteligencia artificial, sin olvidar los usos didácticos y también los trabajos lexicográficos, la compilación de diccionarios, *thesaurus* y la "manipulación" de amplios *corpora* textuales.

La informática distribuida ha creado el terreno indispensable para que cada uno de los usuarios pudiera aprovechar mejor las extensas capacidades ofrecidas por el ordenador en todos sus posibles sectores de aplicación. Esto vale mejor para todas esas disciplinas que en el lenguaje tienen sus bases, y de manera particular para el derecho.

En los años 70 en casi todos los países del mundo así llamado "civilizado" se asistió al desarrollo de grandes sistemas infor-

mativos, con almacenaje de grandes cantidades de datos, pero que eran de estrecha pertenencia de los entes que los crearon y no podían ser utilizados por la mayoría de los usuarios con facilidad.

Además, tales bases de datos eran de exclusivo dominio de los "expertos", quienes procligaban las informaciones a través de procedimientos complejos o con sistemas de búsqueda que estaban caracterizados por ser de élite, es decir, reservados a pocos utilizadores. En las exposiciones y presentaciones de los varios archivos de aquel período se mostraba una supuesta facilidad de consulta, pero, en realidad, existía una notable distancia entre el usuario y el sistema informativo.

En el decenio sucesivo se ha producido el verdadero cambio de calidad, en el sentido de brindar la disponibilidad de las informaciones y de alcanzar un número más grande de usuarios. Esto se obtuvo con la comercialización de computadoras que poco a poco se reducían de tamaño y de precio, mientras que crecían sus capacidades de cálculo, de almacenaje de datos y sobre todo tenían una interface más fácil.

El empuje definitivo fue permitir la comunicación de todos los que utilizan una computadora.

Solamente un ejemplo vale para evidenciar los progresos técnicos: hace veinte años la máquina más difundida en el mundo, la IBM 370/158, tenía 1024 K de memoria central, ocupando grandes espacios con sus discos y sus accesorios. Ahora, una portátil ocupa el espacio de una hoja de papel y tiene normalmente de 2 a 8 M de RAM.

A través de la conexión con la red telemática Internet y la consulta de los bancos de datos que existen en el mundo, cada usuario puede crear archivos personales y analizarlos con su propia computadora.

El Instituto di Linguistica Computazionale ha pasado a través de varias fases de este cambio y, después de haber acumulado un patrimonio de textos memorizados en varias lenguas, ahora pone al alcance del usuario un sistema completo de tratamiento de archivos textuales, de manera que el estudioso pueda llevar a cabo operaciones que antes eran de competencia exclusiva de los informáticos.

En esta comunicación se describen las funciones ofrecidas por el sistema DBT (Data Base Textual), desarrollado por el doctor Picchi, investigador del ILC, a través de algunos ejemplos de aplicación en el sector jurídico.

Las líneas guía que han caracterizado el planteamiento y el desarrollo del sistema, conforme a los procedimientos desarrollados en los años precedentes y las investigaciones actuales, son las siguientes:

- respeto total de las cualidades lexicográficas del material textual a elaborar;
- capacidad de obtener, en tiempo real y de manera interactiva, todas las funcionalidades de un sistema de análisis textual automatizado (búsqueda de palabras en el texto, cálculo de frecuencias, concordancias, *index locorum*, etc.);
- optimización del espacio de disco duro ocupado por los archivos que permiten la consulta *full-text*;
- optimización de los tiempos necesarios para contestar a cualquier tipo de pregunta;
- posibilidad de acceso directo de las funciones descritas al *conjunto* de los textos que se van examinando y no solamente uno a la vez; esto se dice *corpus* y puede ser de gran tamaño;
- facilidad de integración, en el sistema, de otras funciones más complejas para desarrollar investigaciones más especializadas.

Esta modularidad del sistema logra la integración con otros procedimientos que permiten análisis ulteriores como, por ejemplo, la lematización semiautomática, que es la atribución del lema y de su categoría gramatical a la palabra dada, o de redactar diccionarios, glosarios, léxicos de autores, etc.

DBT ha sido concebido como un sistema de interrogación textual que toma en cuenta el problema filológico y adapta la investigación a los particulares fenómenos lingüísticos.

La exposición de sus funciones intenta ofrecer al estudioso un instrumento ágil y dúctil, capaz de adaptarse a las exigencias de análisis más específicos. Agrega además la posibilidad de imprimir las frecuencias y las concordancias del texto. La consulta directa con DBT permitirá al usuario utilizar la gama más amplia de instrumentos para el análisis del texto.

Para dar una demostración de la potencialidad del instrumento y de la investigación posible, consideramos útil presentar una breve descripción de las principales funciones de investigación que están disponibles:

- búsqueda de una determinada palabra en el texto<sup>1</sup>;
- búsqueda de todas las palabras del texto que contienen uno o más caracteres semejantes (tal función puede, por ejemplo, permitir la búsqueda de todas las palabras de un determinado

paradigma, obviando la falta de una lematización del mismo)<sup>2</sup>;

- cálculo del número de veces que aparece cada palabra dentro de un texto o de un conjunto de textos;
- visualización de los contextos de una palabra con diversos modos de presentación (posibilidad de modificar al archivo del contexto, visualizar con modalidad KWIC, lo que significa presentar una única línea con la palabra clave al centro o de otras maneras según le interese al usuario);
- definición de las funciones de búsqueda en el texto de más palabras bases y con operadores lógicos (el operador AND identifica la función por la concurrencia, mientras que OR determina la reunión de ellas en una única unidad de palabras homogéneas);
- búsqueda de una palabra que, mediante una fórmula basada en evaluaciones estadísticas, muestra más estrechamente la conexión de una palabra o de un grupo de palabras seleccionadas por el usuario (*mutual information*);
- conjugar y sincronizar mediante varios mecanismos (diccionarios bilingües, normas, segmento de texto, etc.) un texto determinado con su traducción a otra lengua.

Tiene además otras funciones *ad hoc*, según la necesidad del usuario, como, por

1. Esta aplicación ha sido realizada por el proyecto de traducción al italiano del Digesto, bajo la dirección de Sandro Schipani, profesor de la Universidad de Sassari y de Roma.

2. El proyecto es fruto de una colaboración entre IDG y ASSIA (Associazione di Studi Sociali Latini Americani) y comprende los siguientes textos: 1) Decreto de Argonauri; 2) Bases para un proyecto de Constitución para la República de Venezuela, una e indivisible; 3) Proyecto de Constitución para la República de Venezuela; 4) Mensaje al Congreso Constituyente de Bolivia; 5) Proyecto de Constitución para Bolivia; y 6) Decreto originario del 28 de agosto de 1926.

ejemplo, la posibilidad de tratar diversos alfabetos distintos del latín.

Entre otras cosas que pueden indicarse, respecto de las consultas que pueden realizarse con DBT, encontramos el *Vocabolario Giuridico Italiano* que comprende 107 textos de particular interés para la historia del derecho italiano, que viene desde 1723 hasta 1972, con un total de 2.702.685 palabras. Otro proyecto significativo para la historia del derecho es toda la bibliografía de textos jurídicos antiguos en lengua italiana, desde 1500 hasta 1800. Tal proyecto, llamado BEGA, ha dado como resultado un archivo bibliográfico de fácil consulta que ahora se utiliza con DBT y próximamente mediante la red Internet, de tal manera que pueden consultarse 16.000 títulos que se encuentran en varias bibliotecas italianas y extranjeras, hecho en colaboración con el Instituto de Documentación Jurídica del Consejo Nacional de Investigación (CNR), con sede en Florencia.

Otro archivo básico para la historia del derecho es el que contiene la totalidad de fuentes del derecho romano, con el Digesto, el Codex y las Novellas de Justiniano, tanto en griego como en latín, y al cual se incorporará muy pronto el Epítome Juliano. Estas fuentes se pueden consultar tanto singular como conjuntamente. Para la consulta del Digesto<sup>3</sup> ha sido desarrollada una función particular que permite preguntar tanto por el texto latino como por su traducción italiana, pudiendo recuperarse automáticamente el contexto correspondiente.

Como ejemplo de indagación de un texto en lengua española, existe una función particular del DBT, llamada *mutual information*, con la cual se puede interrogar por un texto mediante un cálculo esta-

dístico que se basa en los valores de frecuencias de aparición de una palabra, proporcionada por el usuario, conjuntamente con otras que aparezcan en el mismo contexto con más frecuencia (4-5-6). Mediante la selección de una palabra se puede obtener una lista de todas las palabras que se encuentran más estrechamente ligadas a la palabra dada. Habiendo obtenido estas palabras, uno puede ver el contexto original del tema en cuestión. El ejemplo de consulta parte de un conjunto de textos memorizados por el proyecto denominado "Léxico constitucional bolivariano", hecho para el estudio del pensamiento político jurídico de Simón Bolívar.

Para la lengua italiana existe la posibilidad de consultar un diccionario conjuntamente con un texto; a través del diccionario y de una serie de reglas definidas sobre bases estadísticas, pueden reconocerse varias partes del texto y proponerse un análisis completo del mismo, comprendiendo el lema y la categoría gramatical.

Para concluir, deseamos subrayar la importancia de la contribución que puede dar la lingüística computacional a varios sectores de la informática jurídica, y si bien hemos presentado ejemplos sobre la historia del derecho, en realidad estamos colaborando en materia tributaria, así como con la legislación vigente. En nuestra actividad falta una colaboración más estrecha con los jueces, que en Italia no recurren al computador a pesar de que en este campo tenemos un archivo que comprende las principales sentencias emitidas por el Tribunal Permanente de los Pueblos y en el que existen muchas referidas a países de América latina.

3 Ver nota 1.