

Este documento está publicado en:

Moreiro González, José Antonio. (1994).
Documentación y Lingüística: conceptos de relación
esenciales. *Ciencias de la Informacion*, 25(4), pp.
202-212

Documentación y lingüística: Conceptos de relación esenciales

José A. Moreiro González

Se reflexiona sobre los conceptos que sirven de enlace tanto en lo teórico como en lo aplicativo a los campos de la Documentación-Información y la lingüística. Su comparación viene razonada a través de las relaciones históricas entre ambas disciplinas. Desde donde se pasa a considerar las intervenciones lingüísticas en la problemática planteada por el análisis semántico de los documentos. Luego se hace una breve descripción de otras relaciones lingüístico-documentales, aplicadas principalmente a la gestión de los sistemas de información automatizados.

The concepts used for the relation of the documentation-information and the linguistics fields, as much as in theory and practice, are thought-out. Their comparison has been realized through historical links between both disciplines. Furthermore, the linguistic interventions on the matter, derived from the semantic analysis of the documents, are considered. A brief description of other linguistic-documentation links, mainly applied to the management of automated information systems, is made herein.

Naturaleza común: razones históricas

Si consideramos que la función del documentalista consiste en intermediar entre los documentos y la información, el almacenamiento y recuperación de esos documentos, encontramos que en estos aparece ya una doble relación con la lingüística:

- Primero, por ser la lengua el vehículo mediante el cual se significan.
- Después, por formar parte del discurso general de la ciencia (del proceso comunicativo humano y científico), cuya expresión común continúa siendo el texto escrito.

Son, por tanto, palabras. Palabra en cuanto constructo abstracto de la realidad referente. Palabra que transmite los conceptos que los documentos abarcan. Y palabra-término por tratar de mensajes especializados. Palabra sometida a las formas y normas de la lengua. Además, su contexto es ser después de otros y antes de los que vendrán. Cada texto científico se sitúa tras un conjunto extenso de investigaciones que le preceden y sin las cuales, por continuidad o por ruptura, no tendría sentido. A su vez, cada texto posibilita, por sucesión o por contraste, el crecimiento del conocimiento humano.

Los documentos científicos, insísimos de mayoritaria expresión textual, se caracterizan por:

- Emitirse con la intención de aumentar el conocimiento.
- Depender, para ser aceptados, del conocimiento previo que posean los receptores.
- Manifiestar la necesaria continuidad de sentido que los haga comprensibles,* por lo que:
 - . Están sometidos a la gramaticalidad.
 - . Se establecen dentro de la pragmática de la comunicación.
 - . Como objetos de comunicación científica presentan una disposición característica, una terminología propia y una particular construcción de las frases.

* Sobre las exigencias de textualidad que deben cumplir los documentos científicos, véase (1); respecto a los requisitos de los textos científicos, véase (2).

La presencia del lenguaje resulta, pues, esencial en el proceso comunicativo de la ciencia. Partimos de la lengua como canal de transmisión de los documentos. Además, la intermediación documental exige trabajar el documento para hacerlo accesible, explicarlo, y hasta gestionarlo. Los documentos son estructuras lingüísticas. Y desde la lengua los procesamos, y representamos su información.

En su desarrollo histórico la ciencia ha sufrido crisis metodológicas coincidentes con sus grandes momentos de crecimiento. Cada vez que esto sucedía se ha reflejado en la comunicación de los estudios científicos. La funcionalidad del discurso de la ciencia no puede comprenderse si no apreciamos la Bibliografía como respuesta a la multiplicación que causó la aparición de la imprenta. Desde que comenzaron a elaborarse las bibliografías, o cualquier cuerpo referencial precedente, el hombre ha tratado de poner orden en el caos de la producción científica de originales. La historia ha conocido un desarrollo simultáneo del orden y de sistemas cada vez más sofisticados de codificación. En el siglo XVII la revolución científica exigió vías más rápidas y eficaces de transmisión, y aparecieron las publicaciones periódicas. La dispersión que su aparición supuso fue contrarrestada por las revistas de resúmenes y los boletines de sumarios. Después, la Revolución Industrial provocó la aparición de nuevos soportes documentales y el crecimiento exponencial de los documentos. Nació entonces la Documentación como respuesta a una información cuyos objetos se diversificaban y dispersaban. Sin embargo, la plenitud de la intervención tecnológica no se dio hasta mediado el siglo XX. Llegaba la Ciencia de la Información, aprovechando la rapidez y fiabilidad de las comunicaciones, para hacer disponibles los mensajes requeridos por encima de las distancias.

Cada una de estas actuaciones buscaba asegurar el acceso a la información existente. Y en cada una de ellas se profundizó en el aprovechamiento de lo que los mensajes ofertaban para representarse y orientar hacia el conocimiento deseado. Primero, la indización de los libros, y la descripción y resumen de las bibliografías; luego, tras la irrupción de las revistas, la aparición periódica de barridos analíticos, que supuso ya la utilización de elementos localizadores y explicativos de una oferta científica en progresiva diferenciación. Históricamente la utilización del lenguaje con fines de acceso y control a la información partió del lenguaje natural. Los conceptos que contenía un libro, y más tarde una revista o colección, se reflejaban en los índices mediante términos, cuyo control sólo dependía del conocimiento y la profesionalidad del autor o del indizador. En la ejecución de los índices podríamos, como mucho, encontrar el reflejo de las primeras clasificaciones científicas, en concreto las derivadas de la Clasificación Enciclopédica de Aristóteles, o las influidas por la División de los Conocimientos de Bacon.

La documentación-información maneja datos y documentos, objetos cuyo atributo es la información. Esta, de naturaleza conceptual, para hacerse evidente debe incorporarse a una representación física, sobre la que actuará el hombre para alcanzar el conocimiento. Toda intervención humana para derivar nuevas formas informativas, manejar mediante la tecnología, almacenar y recuperar en sistemas concretos, se efectúa sobre esta forma física de información. La información fijada lo ha sido mediante la lengua como proceso en los discursos orales y escritos, y es tratada en éstos y también en el resto de discursos a través de la lengua. Cualquier tipo de documento es producto de la capacidad semiótica del hombre para elaborar sistemas de signos. Este carácter de permanencia del documento es lo que diferencia precisamente al discurso documental de otros discursos informativos o lingüísticos. El subsiguiente procesamiento sólo es posible porque la información ha sido previamente grabada. Y es esta condición la que permite almacenar, tratar y difundir los discursos humanos por encima del tiempo y de la distancia.

La fundamentación disciplinar de la documentación, cuando empezaba el siglo presente, consolidó los recursos semánticos como vía regularizada y convenida para organizar y controlar la complicada existencia de los documentos. El intento de Repertorio Bibliográfico Universal no podía ponerse en práctica sin un cuerpo lingüístico reglado que permitiese el entendimiento supranacional.* En el origen de los pensamientos de Otlet se situaba la organización detallada del conocimiento de su época. La enciclopedia global necesitaba de lazos que relacionasen las materias

* Sobre la utilización de la CDU como metodología organizativa del RBU, véase el capítulo "El nacimiento de la Documentación: Dewey, La Fontaine y Otlet", en (3). Los intentos clasificadores anteriores fueron localistas, mientras la CDU supuso un acuerdo internacional, y dio origen a las organizaciones lingüísticas documentales.

ilimitadas que allí se reflejaban. Se reconocía así la capacidad del mensaje documental para sistematizar la información desde categorías lingüísticas, apareciendo los primeros intentos de alcanzar lenguajes para dialogar con los documentos. El propio acceso científico a la lengua, tras la modernización de los estudios filológicos, coincidió en el tiempo, y permitió plantear vías de investigación tendentes a representar con pertinencia el contenido de los documentos. Fue el momento en que se desarrollaron los lenguajes codificados que agrupamos bajo la denominación de sistemas de clasificación.* Surgieron para ordenar los conceptos de las numerosas experiencias, observaciones y entidades de información generadas tras el triunfo de la Revolución Industrial. En razón del momento, se construyeron sobre métodos racionales y positivistas. Lo que conllevaba establecer relaciones entre los conceptos manejados, y estructurarlos jerárquicamente. Se iniciaba así el período en que la CDU y otras clasificaciones documentales se impusieron para representar la información mediante signos convenidos. Estos códigos no se pueden comprender fuera de los lenguajes científicos. Desde el origen de las clasificaciones de utilización documental intervinieron factores lingüísticos: la sustitución de significantes por nomenclaturas codificadas no evita el análisis semántico y su representación mediante signos convenidos. La CDU expresaba con números las abstracciones científicas y la proximidad entre los conceptos por similitud, dependencia o diferencia.

La situación cambió cuando los ordenadores empezaron a intervenir en los procedimientos de transmisión y recepción de mensajes en los años 60. El progreso que esto supuso situó al mundo de las comunicaciones entre la cibernética y los modelos derivados del sistema nervioso humano. De la mano de ingenieros de telecomunicaciones se impuso entonces el sistema de comunicación de datos basado en teorías matemáticas (4). Los mensajes comenzaron a integrarse por conjuntos de datos (bases de datos) legibles por máquinas y organizados lógicamente para su recuperación (5).** Esto supuso un regreso a los lenguajes naturales, si bien controlados. Desde sus primeras aplicaciones los ordenadores atendieron a experimentos lingüísticos, de los que surgieron las llamadas industrias de la lengua. La progresión habida en sus productos ha ofrecido fácilmente una vertiente aplicativa a la actividad documental. A la vez que ha exigido la sistematización de las estructuras y normas lingüísticas para poderlas informatizar. La contribución lingüística a la documentación se ha beneficiado directamente del procesamiento automático del lenguaje, y desde entonces ha producido elementos de gran provecho para la gestión del contenido documental.

Precisamente cuando en esa década se establecía el concepto de information science, se destacó de inmediato su relación con la Lingüística.

Y se hizo tanto desde la perspectiva comunicativa en la que ambas participan (6), como desde su común aportación a la comprensión de los procesos de transferencia de información (7, 8). Los años 60 impusieron la indización desde conceptos lingüísticos y con apoyo informático. Los lenguajes combinatorios han evolucionado desde entonces, primero hacia una gestión automática, y actualmente, apoyándose en la indización automática, se disponen a viabilizar respuestas a preguntas libres de los usuarios. Se evitaría así tanto la ambigüedad y el ruido de las demandas realizadas en lenguaje libre, como la limitación de la recuperación planteada a un nivel determinado de indización. El documento y la demanda se reconocerían en lenguaje libre, mientras que el lenguaje documental encauzaría el entendimiento de ambos. El thesaurus como lista de términos dejará de ser necesario, y se centrará en intermediar comprensivamente la demanda del usuario y el contenido documental. Esta aspiración de facilitar al usuario la recuperación, sin que esta pierda precisión es el reto de las investigaciones aplicadas a la indización automática.

Estamos viendo cómo las demandas que se plantean a las colecciones documentales se apoyan de manera creciente en el tratamiento automático de la lengua natural (9):

- La posibilidad de simular en las máquinas los procesos inteligentes humanos abrió a partir de los años 80 el tratamiento del lenguaje desde la representación informática del conocimiento.

* Inspirados en las estructuras jerárquicas del pensamiento aristotélico-escolástico, y posteriormente del Racionalismo. Esta situación contextualiza el intento de Olet de alcanzar un Repertorio Bibliográfico Universal, que precisaba de un lenguaje de comprensión internacional que contuviese la totalidad de materias tratadas por los documentos.

** La obtención y el procesamiento de los datos comenzaban a superar así las barreras concretas que presentaba al mundo de la comunicación humana, y en especial el referente a la ciencia y a la tecnología.

- Comprender los procesos lingüísticos profundos ha supuesto avances considerables en los productos requeridos para la gestión y representación documental.
- El progreso de la indización automática y de los sistemas expertos se ha vuelto primordial para alcanzar la automatización de los procesos lingüístico-documentales.
- Basado todo ello en la existencia de los archivos digitales que han permitido almacenar grandes corpus de textos (también de imágenes y sonido), con lo que la disponibilidad y tratamiento de la información ha alcanzado posibilidades hasta hace poco impensables.

Funcionalidad lingüística de la documentación

El discurso científico requiere del discurso documental para que su comunicación se haga efectiva y práctica. El documento es un acto transitivo, cuyos contenidos son informaciones que se hacen vanas si no alcanzan a los destinatarios. Precisamente, dentro de ese contexto comunicativo justifica su actuación el documentalista. Su función social consiste en relacionar a emisor y usuarios a través de una organización compleja:

- 1) La documentación se ha hecho necesaria por garantizar ese proceso de comunicación.
- 2) Debido a lo cual su naturaleza consiste, más que en ninguna otra cosa, en participar en el sistema de búsqueda de algún mensaje, emitido por algún autor, con destino a algún usuario.
- 3) Ese mensaje se emite, se recibe, se trata y se refiere a través de la lengua.

Así, pues, el objeto al que atiende la documentación, la información, se hace palpable a través del lenguaje.* Partimos de los documentos en cuanto configuraciones lingüísticas permanentes, y a ellas volvemos tanto en la indización (palabras en cuanto términos) como en el resumen (retextualización de un texto). De tal manera que la disposición de estructuras que trasladamos, los medios analíticos, y la traducción a un producto documental nuevo se realizan por caminos lingüísticos. La mayor preocupación del teórico de la indización estriba en establecer relaciones entre el lenguaje natural y los lenguajes documentales. Aquel como conformador del discurso original. Estos en cuanto traducen un documento a enunciados singulares que hacen posible su recuperación, y que, originados en el lenguaje natural, han sufrido un proceso de normalización previo para constituir una sintaxis particular "en un campo semántico previamente determinado" (11). En el caso del resumen, al pasar del original a su explicación reducida, nos movemos siempre en el lenguaje natural. Vemos pues como el uso de la lengua y de los lenguajes documentales alcanza en la indización y el resumen su mayor relevancia informativa. Y que la fundamentación teórica y metodológica del análisis de contenido se origina en la lingüística. Evidentemente con un objeto y unos objetivos propios.

Cumple a la ciencia de la información actuar como factor pragmático de la comunicación científica. El *discurso de la ciencia* se ha ido componiendo por agregación (por avance, revisión o ruptura) de todos los contenidos aparecidos en los documentos científicos. El conocimiento humano ha ido formando un macrotexto, donde se han reflejado las grandes teorías generales, y la división en ámbitos disciplinares con una conceptualización y metodología propias. El desarrollo del conocimiento en un momento dado es el contexto donde aparecen los productos científicos concretos, como resultado de investigaciones cuyo contenido se fija en documentos. Para que el proceso textual sea completo, el mensaje tiene que ponerse en disposición de consulta. Tiene que llegar al receptor y, luego, ser entendido. La misión primera de la documentación consiste en recoger información (mensajes expresados en cualquier tipo de soporte) allí donde se haya dado, actuar sobre ella con intención analítico-sintética y reofertarla a los ocasionales usuarios. Se establece esta vía como la única posible y razonable en el adecuado trato de los conocimientos científicos. La documentación se aprecia, pues, como transcurso hacia la comunicación de la información. Este es su principal papel: explicar y representar cada discurso científico concreto,

* Estas afirmaciones se sitúan dentro de la postura defendida por García Gutiérrez, quien considera a la documentación como actividad lingüístico-informativa en los ámbitos científico y profesional (10).

como paso obligatorio por el que se sortean las dificultades y se permite a los receptores finales acceder a los datos.

Para realizar esta actividad aplicativa la documentación, además de aprovechar los resultados de la investigación lingüística, tiene a esta como referencia teórica. El documentalista en su intermediación se obliga de una parte a seguir los procesos lingüísticos tanto de comprensión, como de expresión, y por otra a atenerse a categorías de índole documental. Para lo cual trabaja sobre el documento en estas fases, en cada una de las cuales intervienen factores lingüísticos de aplicación directa a la documentación:^{*}

- 1) **Lectura-análisis:** que consiste en identificar y comprender el contenido de los documentos y su distribución.
- 2) **Síntesis:** Mediante la cual se interpreta y simplifica el discurso a las líneas macroestructurales para poder manejarlo.
- 3) **Representación:** Re-expresión de la información fundamental del texto, buscando su comunicabilidad. Su reestructuración sitúa al producto del tratamiento documental en disposición de remediación en el uso social de la información manejada.

En esta propuesta hay una cuestión clave: ¿cómo pasar desde el documento completo a su representación? Es aquí donde se aprecia la necesidad de la lingüística para realizar el tratamiento técnico fundamental de nuestro ámbito. El profesional de la información científica valora las cualidades informativo-comunicativas de un documento dado, y desde su conformación original lleva su contenido hasta cumplir una función referencial. Por ella, y con soluciones semánticas provenientes del propio mensaje del documento, se superan los problemas de acceso a la existencia y datos de los discursos científicos concretos. Esta operación analítica se sirve del lenguaje como método para captar y reofertar la información.

La existencia de la bibliografía primero, y luego de la documentación-información se justifica en avisar a los investigadores de la existencia de unos documentos cuyo contenido les podría interesar. Se quería dar a conocer lo que decían los documentos, su auténtica razón de ser. La solución principal y original consistió en aminorar hasta lo sustancial la información que contenían. El análisis documental debe entenderse desde metodologías que identifiquen los atributos descriptivos, los argumentos y la disposición de los textos partiendo de esquemas lógicos. Para ello precisamos profundizar en conceptos, luego instrumentalizables que "permitan reconocer las estructuras discursivas, e identificar las informaciones substantivas contenidas en ellas".** A este análisis de la estructura informativa profunda atienden dos niveles de descripción documental:

- **La indización:** o identificación de los documentos por conjuntos, de acuerdo con significaciones coincidentes. Sus resultados reflejan en términos relevantes y pertinentes los conceptos que contiene el documento. Sobre esos términos preferentes de descripción se sitúa el eje procesual de nuestra ciencia, pues en torno a ellos se disponen los métodos más aceptados para recuperar la información. Heredera de las técnicas seculares de indización, cuya filosofía ha sido siempre guiar al lector hasta la ubicación exacta de la información que persigue, hoy la recuperación documental es la elección de las obras existentes sobre un asunto científico o general cuya información se adecua a las necesidades del demandante (14). La forma más común de recuperación se establece mediante los descriptores, que se obtienen por aplicación de los lenguajes documentales a esferas concretas del conocimiento. Necesitándose además la combinación de esos elementos a través de las operaciones de álgebra de Boole. Así como de una aplicación por la que se establece el proyecto de recuperación, la estrategia de búsqueda. La adecuación entre los términos que describen fenómenos de un lenguaje especializado y la denominación de los hechos narrados por un texto concreto viene dada por un instrumento del lenguaje documental, el thesaurus, primera norma de pertinencia y relevancia de los descriptores.

* Véase (12).

** La representación consistente tanto del resumen, como de la indización, en cuanto traducción de contenidos, depende de la "identificación también consistente de la información más relevante contenida en un texto dado" (13).

Los sistemas de clasificación e indización controlan los términos y signos de representación, que organizan e identifican los documentos por su contenido. Los términos controlados se aplican para caracterizar contenidos en los documentos. A través suyo se realiza, posteriormente, la recuperación, y son el elemento de acuerdo de los distintos profesionales y los usuarios para denominar las materias documentales. Ya vimos antes cómo la aspiración deseable sería lograr tesauros inteligentes que, además de agrupar y relacionar descriptores, proporcionen respuestas (desde un almacenamiento de información) a preguntas concretas.

- Y el **resumen** que consigue la reducción del mensaje mediante la expresión más completa, rica y explicativa del documento, utilizando el discurso en desarrollo pleno. Este nivel descriptivo, tras identificar las estructuras semánticas fundamentales, las explica conforme a los postulados del lenguaje natural. El problema esencial de las traducciones que se efectúan a través de la indización y del resumen consiste en explicar las reducciones informativas de tal forma que el documento pueda reestructurarse en sus líneas semánticas principales desde la serie de sus descriptores o desde su resumen.

Los servicios de resumen y de traducción automática coinciden en representar un texto en otro texto, ya sea reducido a sus líneas macroestructurales, o expresado en una lengua diferente. En este caso facilitan la consulta de los originales, y apoyan el tratamiento a documentos escritos en otros idiomas: desde el deseo de romper la incomunicación creada por la diversidad de lenguas existentes, se ha establecido la traducción automática como el objetivo prioritario del procesamiento informático del lenguaje (15). Hay que considerar que la traducción automática es un servicio más de los centros de documentación, por el que se busca facilitar el acceso pleno a los originales y utilizar la lengua propia para entrar en las bases de datos sin que el idioma nos imponga limitaciones. Además, los logros hacia la traducción automática, y el conocimiento de las herramientas que lo permitan, serán aplicables directamente al conocimiento necesario para resumir mecánicamente.* La presencia del ordenador en el tratamiento de los documentos-textos exigió profundizar en el conocimiento de las estructuras de estos, y en el de las estrategias que se precisan tanto para su conformación, como para su entendimiento. Era obligado sistematizar los elementos de la lengua si se deseaba su informatización. En este punto, existe casi una exacta coincidencia en los modelos buscados para la representación documental automatizada con aquellos usados en la traducción automática. Ambos han partido del análisis del documento íntegro para alcanzar:

- Su análisis y comprensión.
- La representación lógica del conocimiento (IA). La IA aparece en ambos casos como necesaria para enfrentarse al análisis semántico y contextual de los textos.
- El tratamiento automático de la lengua, dirigido a los Sistemas de indización y resumen, o a los sistemas de traducción y sistemas terminológicos.

La solución a los problemas planteados por el tratamiento automático del lenguaje obligó a hacer aproximaciones a este de acuerdo con textos completos, no por frases aisladas. Los modelos buscados por los especialistas en documentación automatizada corrieron muy próximos a los usados en la traducción mecánica y planteaban el obligatorio análisis de los textos íntegros. Esta necesidad fomentó en parte desde la Documentación-Información el desarrollo de la ciencia del texto, y recibió en reciprocidad solución a los problemas semánticos de la intermediación documental.

Información y lingüística, una relación proyectiva

Hemos visto como en el análisis documental se encuentra la asociación más estrecha de nuestro campo con la Lingüística. Han ido apareciendo también otras relaciones donde ambos sectores se entrecruzan. Actualmente se advierten en esta colaboración nuevas posibilidades destinadas a un más eficaz manejo de la información. A la indudable contribución, ya tradicional, de los sistemas de indización y

* Véase (16).

resumen, han venido a sumarse otros productos lingüísticos, siempre en la persecución de disponibilidad informativa rápida y satisfactoria. De tal manera que el estudio de los lenguajes naturales se ha convertido en un factor indispensable para entender el procesamiento de la información en la ciencia de la información.

Mediante la lengua se expresan los originales, se comunican emisor y receptores, y se relacionan ambos con el analista y difusor documentales. Además, de la lengua derivan las herramientas que utilizan los sistemas de información para almacenar, organizar y recuperar los documentos. Para procesar la lengua y los documentos el ordenador se ha visto obligado a distinguir las estructuras gramaticales, solucionar las situaciones de sinonimia, hacer inferencias desde el conocimiento del mundo, y diferenciar la disposición y orden de los textos. Por su parte, la inteligencia artificial persigue intercambiar informaciones entre el hombre y la máquina mediante el lenguaje natural, sin usar lenguajes intermedios. De ambas posibilidades surgen al momento atribuciones a la traducción y análisis automatizados, a la composición mecánica de textos, y a los analizadores y sintetizadores de voz y escritura. Procesamiento automático de los textos que, como antes comentábamos, requiere el conocimiento de sus estructuras y la explicación de las múltiples operaciones y normas que se precisan para su conformación y comprensión. Para tratar la información de los documentos, tanto humana como automáticamente, debemos conocer cómo se organizan y qué función cumplen las estructuras que los conforman.*

El análisis documental se somete a regulaciones de carácter lingüístico y retórico, así como a operaciones de abstracción y sumarización. Por lo que tiene gran importancia considerar la intervención estratégica del conocimiento. La comprensión de los contenidos implica la actuación simultánea de varias estructuras cognitivas para resolver los problemas de procesamiento y representación textual. Los distintos niveles que actúan coordinados son los siguientes:

- 1) Proceso de percepción de los datos que aportan los términos y las proposiciones.
- 2) Sistema de conocimiento general interactivo, adquirido mediante la experiencia personal, que se define a través del:
 - 2.1) Conocimiento léxico.
 - 2.2) Estructurado en marcos.
 - 2.3) Junto al aporte volitivo y personal del receptor: fines, deseos, intereses, intenciones y emociones.
- 3) Modelo de situación para representar la memoria del documento concreto:
 - 3.1) Características de la expresión concreta.
 - 3.2) Macroestructura y superestructura.
 - 3.3) Coherencia de la estructura proposicional.

Se requiere la explicación de todo ello, su modelización y procesamiento para avanzar en las técnicas documentales de descripción y recuperación. La producción de índices y de textos referenciales demuestra un modo de competencia lingüístico-gramatical. A través de una estructura y siguiendo unos objetivos determinados representamos un discurso con especiales características semántico-pragmáticas. Que, además, surge como consecuencia de los propósitos comunicativos de la ciencia, y de las intenciones y motivaciones científicas del documentalista que busca una interacción con los posibles destinatarios.

Hemos visto antes cómo las operaciones de indización y resumen, fundamentales para el análisis y gestión de la información, se basan principalmente en la lengua. Aunque su elaboración se ha hecho

* La organización estructural, los medios analíticos, y la representación de las referencias documentales se realizan por medios lingüísticos, por lo que no debe extrañarnos el estudio del análisis de contenido partiendo de la lingüística del texto. Esta atiende al nivel semántico de las palabras, de las oraciones y del texto, sobre todo a la globalidad que este concede a los documentos. Además, esta teoría lingüística se orienta siempre hacia la comunicación. Por lo que nos alcanza bases teóricas para encuadrar las operaciones descriptivas del documento.

tradicional en la intermediación informativa, tienen un claro futuro desde la perspectiva del procesamiento automático de la lengua. Sin embargo, la lengua comienza a estar presente en las unidades de información mediante otros medios destinados a mejorar el tratamiento de los documentos, y a apoyar las tareas de los documentalistas y los servicios que se prestan a los usuarios (17). Estos sistemas se encuentran ampliamente repercutidos por lo que se ha denominado *industrias de la lengua* (18).

Para automatizar cualquier proceso de reconocimiento, reducción y representación de la información se precisan unos instrumentos lingüísticos: los analizadores morfológico-sintácticos, semánticos, y pragmáticos. Sin ellos no es posible la localización de palabras vacías. Ni se pueden formar los índices tipo KWIC, ni estudiar la incidencia de los términos en los títulos, los resúmenes, e incluso el texto íntegro. Son los instrumentos que permiten seguir la cohesión y coherencia de los textos, y navegar por las redes semánticas.

Otra ayuda muy importante para apoyar el tratamiento de la información son los sistemas de reconocimiento de voz y de caracteres, que facilitan la entrada y salida de la información. Los sistemas de reconocimiento de caracteres (de tipo óptico especialmente en la actualidad) nos conceden facilidad para almacenar en el ordenador textos escritos tipográficamente. Con una rápida progresión hacia una mayor exactitud lectora. Junto a ellos, las técnicas de reconocimiento de voz quieren facilitar la grabación y gestión oral de los datos. Lo que supondría obtener rapidez y naturalidad en las actividades técnicas de los sistemas de información, comunes en cualquier operación en que intervengan la comunicación y la información.

Contribución destacadísima es la que efectúan los sistemas terminológicos, que fijan los conceptos de cada ámbito del conocimiento. Los términos sirven para ordenar el conocimiento a través de la expresión conceptual del pensamiento, a la vez que difunden ese conocimiento como elementos de los lenguajes documentales.* Terminología y clasificación científica fueron fruto de la necesidad de normalizar el lenguaje para que los investigadores pudieran entenderse en el siglo pasado. El reflejo conceptual de los documentos sobre el que son recuperados se realiza a través de la indización, cuya base son los sistemas terminológicos. La normalización terminológica supone la corroboración de una palabra como descriptor. Los elementos con los que se conforman los tesauros deben estar refrendados como términos preferentes en un campo científico o de la actividad humana. Esa es una de las misiones que deben cumplir los sistemas terminológicos respecto a la Documentación: llevar las palabras desde los léxicos generales hasta los científico-técnicos, y desde estos contribuir a la confirmación de los descriptores en los tesauros.

En el procesamiento automático, la necesidad de que el ordenador comprenda el significado de las palabras, de las oraciones y de los textos, debe hacerle capaz de ponerse en situación de razonar, para así analizar los tres niveles lingüísticos: morfológico-sintáctico, semántico y pragmático.

La existencia de diccionarios y bases de datos terminológicas que administren la información precisa para el tratamiento de los textos. Se necesita que los elementos del lenguaje sean descritos profundamente si se quiere su aplicación real. Estos diccionarios recogerían el estado del asber común que, de esta forma, podría trasvasarse a los ordenadores. El valor contextual de las palabras está siempre presente en el procesamiento automático de los textos que forman los documentos.

La importancia de la lexicografía y de la terminología deriva de su testimonio de fidedignidad. Para asegurar la comunicación objetiva, los lenguajes documentales han exigido la precisión de los términos sinónimos, polisémicos y homónimos. Desajustes entre significante y significado cuya fijación se debe perseguir en los lenguajes combinatorios para asegurar la univocidad de los descriptores y la coincidencia entre la apelación y la petición informativas.

La posibilidad de combinar distintos tipos de datos (hipertextos) y los archivos en soporte electrónico han posibilitado en los últimos años perspectivas amplísimas en el almacenamiento y recuperación de información. Los soportes láser, la ampliación de memoria de los discos y disquetes de ordenador, las cintas, (...) han modificado el concepto de acceso a la información y su procesamiento. A la par, la aspiración a la globalidad del acceso ha sido permitida por el hipertexto, nuevo modo de desplazamiento asociativo por la información con perfiles personales de búsqueda.

* Véase (19).

En estos últimos años se empieza a notar una demanda de instrumentos lingüísticos para ser tratados automáticamente. Los vacíos en la oferta de esos recursos que las empresas observan deben ser rellenados por los investigadores. En sus funciones recae ofertar corpus documentales y proyectar interfaces en lengua natural, sistemas expertos e inteligencia artificial.

Conclusiones

La lengua es un instrumento de gran relevancia en las actividades documentales. Mediante ella se explora y explota la información. Lejos de la lingüística no se pueden explicar las funciones fundamentales que la ciencia de la información efectúa sobre el análisis semántico de los documentos. Muchos de los avances que se dan en su conocimiento se suceden en aplicaciones para los sistemas documentales. La tradicional aportación de la lengua al análisis documental se ve actualmente superada por otras ayudas lingüísticas a la gestión de la información y de los sistemas de información.

Por encima de la aportación aplicativa, la lingüística aporta conceptos desde los cuales comprender la naturaleza de la Documentación, pues la condición de esta y muchos de los métodos de sus procesos se refieren inexcusablemente a la lengua.

El profesional, y el que se prepara para serlo, deben conocer competencialmente los factores lingüísticos que intervienen en el tratamiento de la información. Desde ese nivel de conocimiento estará capacitado para seguir las aplicaciones continuas desarrolladas por la investigación lingüística.

El acceso a la información documental y su procesamiento no puede realizarse sin el conocimiento léxico, lingüístico (sintáctico-semántico) y contextual. Además, en los trabajos científicos, está presente el conocimiento metodológico o de género. Por lo que al Documentalista se le exigen estos niveles competenciales, de clara relación con la lengua:

- 1) Competencia para la actividad volitiva de informar y para lograr la situación comunicativa: expresión, descripciones, referencias y elementos de relación.
- 2) Competencia gramatical y sobre la estructura textual.
- 3) Competencia cognoscitiva sobre el estado de cosas y su representación mental.
- 4) Competencia sobre la situación social en que el texto se da.
- 5) Competencia en el manejo de sistemas automatizados.

Nos seguimos enfrentando al problema superviviente de las grandes colecciones de textos. Entre los que caben mensajes muy variados. Para enfrentarnos a su tratamiento necesitamos potentes métodos de análisis, no limitados a áreas concretas del saber. Nos estará permitido cuando alcancemos la capacidad de analizar con flexibilidad las estructuras textuales, y un mejor conocimiento de las estrategias de reconocimiento y reducción informativas, de cara a una representación más pertinente de la información documental. La lengua, basándose en instrumentos informáticos más completos, tiene un largo camino por delante en su colaboración con la Documentación.

Por lo que a la universidad respecta, los planes de estudio deberán realizarse con la suficiente flexibilidad que permita a los futuros documentalistas integrar su carrera con los perfiles de conocimiento necesarios para enfrentarse a tareas cuyos componentes son básicamente interdisciplinares.

Referencias

- 1) Acosta, L. Cuestiones de Lingüística Textual. Salamanca, Universidad de Salamanca, 1982. p. 29-31.
- 2) Bloomfield, Leonard. Aspectos lingüísticos de la ciencia. Madrid, 1973, p.14.
- 3) Moreiro, J. A. Introducción Bibliográfica y Conceptual al Estudio evolutivo de la Documentación. Barcelona, P.P.U., 1990. p. 65-81.
- 4) Shanon, T., y W. WEAVER. Teoría matemática de la comunicación. Madrid, Forja, 1981.
- 5) Rispa Márquez, R. La revolución de la información. Madrid: Salvat, 1982. p. 38.
- 6) Weisman, H. M. Information and the Discipline of Communication Sciences, *In Proceedings of the American Documentation Institute*. Washington, Thompson Book, 1967. V. 4. p. 8-12.

- 7) Boroko, H. Information Science: What is it?. *American Documentation* 19:3-5, 1968.
- 8) Haynes, R. M. Information Science in Librarianship. *Libri* 19:216-236, 1969.
- 9) Warney-Amy, J. Natural Language Processing. *Annual Review of Information Science and Technology* 22:79-108, 1984.
- 10) García Gutiérrez, A. Connotaciones lingüísticas para una teoría de la Documentación. *Revista Brasileira de Biblioteconomia e Documentação* (São Paulo) 21(1-2):10, 1988.
- 11) Cunha, Isabel M. R. F. Análise documentaria, En SMIT, J. (coord.). Análise documentaria: a análise da síntese. Brasília, IBICT, 1987. p. 46.
- 12) Kobashi, N. Y. Análise documentária. Considerações sobre un modelo lógico-semántico, En Grupo Temma. Análise documentaria. Considerações teóricas e experiências. São Paulo, FEBAB, 1989. pp.47-48.
- 13) Kobashi, Nair Y. *Op. cit.* p. 50.
- 14) Vickery, B. C. *Techniques of information retrieval*. London: Butterworths, 1970. pp. 154-155.
- 15) Díez Carrera, C. La traducción automática. *Signatura* (España-Ministerio de Cultura), 3:18-23, 1993.
- 16) Moreiro González, J. A. El procesamiento automatizado de los resúmenes. En su Aplicación de las Ciencias del Texto al resumen Documental. Madrid, B.O.E.-Universidad Carlos III, 1993. p.p.172-179.
- 17) Díez Cabrera, C. Productos lingüísticos para los centros de información. *Revista Española de Documentación Científica* 16: 137-146, 1993.
- 18) Vidal Beneito, J. (coord.). *Las industrias de la lengua*. Salamanca, Fundación G. Sánchez Ruipérez, 1991.
- 19) Galinsky, C. Terminología y sociedad moderna: el papel de INFOTERM. En Vidal Beneito, J. (coord.). *Las industrias de la lengua*. Salamanca, Fundación G. Sánchez Ruipérez, 1991.

Recibido: 10 de noviembre de 1994.

Aprobado: 12 de noviembre de 1994.

José Antonio Moreiro González
 Departamento de Biblioteconomía y Documentación
 Universidad Carlos III de Madrid
 Calle Madrid 126 -28903 Getafe
 España

Servicios y Procesos de Información

Taller de políticas y técnicas para la selección y adquisición en bibliotecas y otras instituciones de información

Del 15 al 19 de mayo de 1995.

Servicios de información especializada. Cartera de proyectos

Del 5 al 9 de junio de 1995.

Bibliometría e informetría

Del 10 al 14 de abril de 1995.

Los mejores profesores

Amplias experiencias

Temas de actualidad



**Servicios y Procesos
de Información**