

# Evolución paralela de los Lenguajes Documentales y la Terminología

José Antonio Moreira González

Universidad Carlos III de Madrid. [jamore@bib.uc3m.es](mailto:jamore@bib.uc3m.es)

## Resumen

*Se relacionan Terminología y Documentación desde la perspectiva común de servir a la representación de los conceptos científicos. En especial se analiza la evolución de ambas a la par que la teoría científica básica de origen positivista, y la evolución habida tras la II Guerra Mundial, con el posible paralelismo existente entre la recuperación de la información y la terminología, en especial mediante las aportaciones de Cutter, Dewey, Lafontaine y Otlet, y Bush. Se atiende a la función de la terminología en los nuevos lenguajes documentales: tesauros automáticos y Topic maps.*

**Palabras clave:** Ciencia de la información, Historia, Recuperación de la información, Representación de la información, Terminología, Tesauros automáticos, Topic Maps.

## Abstract

*Terminology and Information Science are related from the common perspective of being useful to the representation of scientific concepts. Mainly the evolution of both disciplines is analyzed as well as the basic scientific theory of positivist origin. Likewise the evolution after the II World War is considered, besides the parallelism between the information retrieval and the terminology, in special by means of the contributions of Cutter, Dewey, Lafontaine and Otlet, and Bush. It is pointed out the function of the terminology in the new indexing languages: automatic thesauri and Topic maps.*

**Keywords:** Automatic thesauri, History, Information representation, Information retrieval, Information Science, Terminology, Topic Maps.

## 1 La organización de la información ante el concepto tradicional y actual de terminología

Aunque no pueda considerarse a la Terminología como la única especialidad lingüística que contribuye al análisis y representación de los contenidos documentales, su intervención es

decisiva, pues actúa como protagonista en los demás procesos lingüístico-documentales y es, desde luego, la especialidad que ofrece un paralelismo mayor con los Lenguajes Documentales.

De forma especial, la organización de la información se afana por buscar soluciones pragmáticas a las nuevas necesidades y planteamientos de la comunicación científica aparecidos como consecuencia de la velocidad alcanzada por los procesos de edición y acceso, que ha llevado a cambiar los planteamientos y prácticas hasta ahora acostumbrados. Los instrumentos usuales de organización de la información han participado de unos fundamentos generales, entre los que estaban: pretender una diseminación de aspiración global, partir de unas formas apriorísticas de representación basadas en categorías generales, suponer una esencia universal en la constitución de los objetos y fenómenos, y “procurar un lenguaje depurado de las impurezas e imperfecciones de los discursos de las ciencias humanas para las que se buscaba el modelo de las ciencias más formalizadas” (Pérez Hernández).

De alguna manera estos conceptos reflejan la tendencia iniciada por Wüster (1998, p. 48) al conceder las bases científicas de la terminología, dentro de un razonamiento en el que subyacían los planteamientos neopositivistas de acercarse al léxico como si sólo se tratase de una nomenclatura o taxonomía, y de buscar, por encima de todo, objetividad, claridad y precisión lingüísticas. Precisamente, la aspiración a un lenguaje unívoco y homogéneo, que fuese de aceptación y aplicación universales se planteó entonces desde unas premisas ajenas al lenguaje natural de uso común. Al centrarse en la regulación metódica y el potencial clasificador de los lenguajes, se postergó su cometido principal de articular la comunicación (Sager, 1993). La propuesta de Wüster es razonable si pensamos que quiso formalizar los lenguajes propios de las materias politécnicas. Pero no sucede así cuando se consideran los rasgos del lenguaje utilizado en muchos de los dominios sociales o humanísticos, que propenden a mezclarse con el lenguaje natural común y cuyo léxico no se organiza fácilmente en taxonomías (Felber; Pitch, 1984).

Tabla 1. Concepto de información

<b>Positivista</b>	<b>Postmoderno</b>
Formas apriorísticas de representación basadas en categorías universales	Categorías funcionales
Preponderancia de las relaciones jerárquicas (Taxonomía)	Preponderancia de las relaciones asociativas <sup>1</sup>
Documento como soporte del conocimiento	Primacía de la función comunicativa del documento
Información vista desde la producción o desde la recepción	Información vista desde la relación producción-recepción del mensaje
Función pedagógica de la Documentación	El usuario como sujeto de la interpretación
Sistemas de información de carácter explicativo	Sistemas de información como instrumentos del proceso de mediación

<sup>1</sup> Para algunos autores, la escritura ordinaria es, por lo general, secuencial porque imita al habla que es secuencial. Además, los libros se leen de forma secuencial. Por el contrario, estos mismos autores piensan que la estructura de las ideas no es secuencial, ya que nuestra mente funciona por asociación. Defienden, entre otros, este tipo de organización del pensamiento Wittgenstein (Wittgenstein, Ludwig.- *Tractatus logico-philosophicus*. Madrid: Alianza Editorial, 1982), Foucault (Foucault, Michel.- *L'Ordre du discours*. Paris: Gallimard, 1971), Barthes (Barthes, Roland.- *Le plaisir du texte*. Paris: Seuil, 1973).

<b>Positivista</b>	<b>Postmoderno</b>
Tecnología cuyas reglas debe de observar el usuario	Tecnología para fomentar el ejercicio de la creatividad
Difusión de afán global	Relativización por el contexto o situación del discurso
Búsqueda prioritaria de la precisión del lenguaje	El lenguaje considerado a partir de su función comunicativa
Lenguaje depurado de las imperfecciones de los discursos de las Humanidades	Utilización de toda la riqueza expresiva del lenguaje
Se buscaba el modelo de las ciencias más formalizadas	Cada documento contiene un modelo léxico
Sustantivo como forma de representación privilegiada <sup>2</sup>	Léxico más aproximado al lenguaje natural

Los términos científico-técnicos cumplen, pues, dos funciones, la de representación conceptual, y la de transmisión de los conocimientos especializados. Benveniste (1989) observó que la historia de la ciencia está asociada a la constitución de una terminología propia, de ahí que el propio científico haya actuado como seleccionador de los contenidos que le interesaban. Así pues, la idea de Benveniste de un vocabulario analítico de la realidad no es otra cosa que la terminología de una especialidad que funciona como herramienta de comprensión y desarrollo. Lo corroboran históricamente los índices de libros, el origen de cuya pertinencia se debe a que están elaborados por los mismos autores del texto, así como los sistemas de clasificación de los fondos, ya desde el Museo de Alejandría. Podemos afirmar que los índices tradicionales establecen una red semántica a partir de los conceptos de los que trata un documento, mediante la participación de (Bonura, 1994):

- Conceptos, cuya designación puede hacerse por más de una palabra o término.
- Asociaciones entre conceptos (véase, véase también; subordinaciones,...).
- Apariciones de los conceptos (casos u ocurrencias): recursos de información hacia los que apuntan las referencias.

De manera que, para controlar la dispersión y facilitar la comunicación entre los expertos, es vital que se controle el lenguaje. Sin embargo, no podemos limitarnos a este nivel mínimo. Precisamente, la Teoría general de la Terminología ha recibido críticas por su forma de considerar las cuestiones externas al mundo del lenguaje. Es una consecuencia inmediata de los planteamientos científicos actuales que muestran una clara desconfianza hacia la rigidez categórica proveniente del razonamiento positivista, pues se prefiere un análisis más pragmático que parta de la visión del observador y que tienda a conjuntar los planos hasta ahora separados del sujeto y del objeto, evitando así la segregación artificial del discurso científico y del sentido común (Harvey, 1994).

Las propuestas actuales prefieren destacar como primordial la función comunicativa de la lengua, desplazando la intensidad puesta, hasta ahora, en el carácter representativo. En consecuencia se observa un alejamiento de los intentos de fijar los términos como referentes

<sup>2</sup> La usada para simbolizar la esencia de las cosas. Cabré, M<sup>a</sup> T.- *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártica-Empúries, 1993.

ideales y homogéneos, estrictamente desambiguados y monosémicos, y al servicio preferente de la comunicación entre expertos (Krieger; Bocorny, 2004).

La Terminología clásica no responde, por lo tanto, a las necesidades de tratar la heterogeneidad y la multiplicidad de puntos de vista respecto a la producción, el uso y el consumo informativos. De ahí que la Socioterminología se interese por las relaciones semánticas, buscando entender adecuadamente la estructura interna del léxico. De forma que las definiciones de carácter dispositivo son desplazadas por descripciones del significado de las palabras, otorgándoles una condición más cambiante, al considerarse las metáforas y los contextos históricos de uso, al tiempo que se favorecen las dudas sobre el sentido monosémico (con lo que se rehabilita consecuentemente la polisemia), y sobre la variación vista como una "perturbación de la unidad lingüística" (Temmerman, 2001).

Si aceptamos que pocos conceptos existen de forma objetiva, veremos que la definición de concepto en la terminología tradicional era restrictiva. Lo que obligó a la Teoría Sociocognitiva a buscar una descripción realista de los significados de los términos tal como aparecen en los textos. Este hecho supuso un serio perjuicio para la visión centralizadora pretendida por los estándares. El contenido pasó a delimitarse desde el texto en el que está insertado el término, por lo que un concepto no es universal, ni inmutable, si no que está elaborado "a través del conocimiento de mundo, de los valores sintácticos, del entendimiento lingüístico-semántico por el que sabe del estado de la cuestión de su objeto y utiliza con exactitud las representaciones de sentido entre los conceptos del texto" (Moreiro González, 1993). De manera que la información trasladada por una definición varía de acuerdo con el tipo de conocimiento tanto del emisor como del receptor. Por la funcionalidad, por la flexibilidad de unos modelos cognitivos en alteración constante, y por la diversidad de los procesos de categorización, suponía pasar a admitir la presencia de la sinonimia y la polisemia (Lara, 2006).

## **2 El uso de la terminología en recuperación de la información**

Únicamente la asignación de enlaces aclaratorios concederá a los lenguajes documentales la posibilidad de cumplir su función comunicativa, interpretativa y recuperadora. Precisamente aquí reside la causa de que la terminología siga siendo una referencia a la hora de ofertar los signos que se corresponden con las diversas posibilidades de interpretación, como luego veremos al hablar de las últimas tendencias en los lenguajes de recuperación y representación. Por lo que respecta a los instrumentos de organización del conocimiento hemos de aceptar que es comprobable la fragilidad que muestran los lenguajes documentales ya que carecen de la deseable doble articulación hacia y desde los usuarios, lo que les hace perder contacto con la realidad (Gonzalo; García Yebra, 2004).

La circulación social del conocimiento ha exigido formalizaciones informativas, de las que se han beneficiado su adquisición, desarrollo y transferencia. Desde una perspectiva contemporánea, en Estados Unidos, el acceso a la información por materias abrió la trayectoria espectacular que tendrían los lenguajes documentales en el siglo XX. En 1876 se dieron a conocer las obras básicas Clasificación decimal de Melvin Dewey (1979) y las *Rules for a dictionary catalog* de Charles A. Cutter (1962). La primera analizó la totalidad del conocimiento humano y lo incluyó dentro de unas divisiones decimales, siguiendo la idea de aspirar a una expresión universalmente válida que diese una visión general de los

conocimientos<sup>3</sup>. Se trata, pues, de una graduación de disciplinas científicas cuya secuencia se basa en el nivel más amplio de la clasificación del saber humano de Francis Bacon (1980) y en niveles medios según los conceptos del siglo XIX, hasta llegar a niveles específicos (Cutter, 1972), por lo que sus características más notables son la simplicidad, la ingenuidad y la adaptabilidad de sus anotaciones<sup>4</sup>; mientras la segunda lo hacía de alguna forma, para los lenguajes controlados, mediante las listas de Encabezamientos de materia, que definió dentro de un carácter precoordinado, estructura asociativa y control de vocabulario de aplicación específica. Sin duda la amigabilidad de uso para el usuario, frente a la rigidez arbórea de los sistemas clasificatorios, dio a los encabezamientos una proyección hacia los lenguajes controlados superior a la que tendrían las clasificaciones.

En un contacto más directo con los presupuestos de los que partió la tendencia documental española, la obra de los bibliógrafos La Fontaine y Otlet pasa por ser un hito en la organización del conocimiento, y enlaza directamente con las teorías positivistas<sup>5</sup>. Una de sus tareas fundamentales fue la organización y puesta en marcha del *Instituto Internacional de Bibliografía* para crear el *Repertorio Bibliográfico Universal* (RBU) cuyo desarrollo exigía contar con la Clasificación Decimal Universal (CDU), pues sin ella no habría posibilidad de cooperación bibliográfica internacional<sup>6</sup>. A principios de 1896 La Fontaine y Otlet habían traducido y normalizado parte de la Clasificación Decimal de Dewey<sup>7</sup>. En este lenguaje los números expresaban abstracciones científicas y permitían circular entre los conceptos extraídos del contenido de los originales, para responder a búsquedas concretas<sup>8</sup>. Como herramienta metodológica la *CDU* estableció enlaces semánticos entre los conceptos del *Repertorio*, relacionándolos por jerarquía, similitud o diferencia. De manera que la *CDU* presenta más características de taxonomía, que de asociación.

---

<sup>3</sup> En la actualidad sigue utilizándose en muchas bibliotecas. Incluso es el origen de la *CDU*, tras la adaptación que de ella hicieron Otlet y La Fontaine. Incluso anticipó sus propuestas a la Clasificación de la Biblioteca del Congreso.

<sup>4</sup> Puede considerarse la *CDU* como una versión de la Clasificación Decimal de Dewey para las bibliotecas, ya que en lo esencial su sistema es el mismo. El conocimiento se encuadra en diez clases con los símbolos gráficos universalmente más conocidos, los números (del 0 al 9). Cada clase de contenido se divide a su vez en diez subclases. Así se forman, por subdivisiones jerárquicas, los índices donde cada sucesiva división de un tema corresponde a una subsiguiente anotación con un nuevo dígito.

<sup>5</sup> Para conocer la aportación de Comte y Spencer al positivismo, véase.- Ducasée, P.- La *synthèse positiviste: Comte et Spencer*, en *Revue de Synthèse*, 1950, 26: 154-163.

<sup>6</sup> Aplicar la Documentación a la producción intelectual universal exigía una cooperación internacional que permitiese regular los intercambios intelectuales dentro de un sistema organizado en el que se comprometiese la administración pública, como una función más de gobierno, a la manera en que lo eran la enseñanza, las obras públicas o la sanidad. Otlet, P.- *L'organisation rationnelle de l'information et de la documentation en matière économique*, en *Bulletin de l'IIB*, 1905, 1-6: 18.

<sup>7</sup> Se preocuparon también de explicar la naturaleza de esta Clasificación. *Organisation internationale de la Bibliographie scientifique*. Bruxelles: OIB, 1895. Siguieron publicando tablas referentes a otras disciplinas científicas, o simplemente ofrecían la traducción de la edición americana de 1894.

<sup>8</sup> Rayward considera a ambas como las funciones básicas de los ordenadores, en su *Reestructuring...*, en *FID.- International Organisation and Communication of Knowledge: Selected Essays of Paul Otlet*. Amsterdam: Elsevier, 1990: 61.

Otro factor decisivo partió en la posguerra mundial de Vannevar Bush, para quien el problema principal estaba en la inadecuada forma de almacenar y clasificar la información y por ello se afanaba en encontrar un sistema de procesamiento que resultara efectivo. En su artículo de carácter prospectivo “*As we may think*” fijó los argumentos para entender lo que varias décadas después sería conocida como *Sociedad de la Información*. Bush comprendió que la estructura secuencial de los documentos era la causante de que los métodos de su tiempo (taxonomías alfabéticas o numéricas que disponían los conceptos en clase-subclase) fueran incapaces de procesar adecuadamente grandes cantidades de información, pues “la mente humana no trabaja de esa manera, sino que opera por medio de asociaciones”<sup>9</sup>. Vio que los sistemas jerárquicos quedaban superados por la asociación de conceptos, imitando el modo en que las personas piensan. Incluso llegó a diseñar una máquina informativa, el Memex, a imitación del funcionamiento asociativo del cerebro humano<sup>10</sup>, por lo que no sólo almacenaría informaciones, si no que además las combinaría en hipertexto, poniendo remedio a la recuperación de un conocimiento imposible de abordar linealmente debido a su desarrollo incontenible. Bush consideró que los sistemas de indización eran de gran artificialidad para acceder al conocimiento, pues los sistemas de clasificación o categóricos por materias seguían la idea de la jerarquía lógica y su rigidez era lo más alejado del funcionamiento asociativo de la mente. La novedad que introdujo el *Memex* radica en su sistema flexible de “índice por asociación” a través de conexiones, y en la posibilidad de agrupar distintos tipos de información que puede ser compartida por varios usuarios. Era evidente que las categorías tradicionales de la Biblioteconomía resultaban insuficientes para tratar el contenido de los documentos, pues las clasificaciones universales contribuían poco a la circulación efectiva de los contenidos registrados. Con esta propuesta Bush se internó en los índices asociativos y predijo la implantación de los sistemas de bases de datos, hipertextos e hipermedia<sup>11</sup>.

### 3 Organización de la información, lenguajes documentales y terminología

Volviendo a Wüster (1998), hemos de aceptar que influyó en la creación del Comité Técnico 37 (TC 37) de la ISA (*International Standardization Association*), fundada en 1926, y luego ISO (*International Organization for Standardization*), cuyo objetivo se situaba en unificar los métodos de trabajo y las terminologías. Interrumpidos sus trabajos durante la II guerra mundial, se reorganizaron en los años cincuenta. Sin embargo, no fue hasta 1979 cuando se hizo evidente la relación entre terminología y normalización, fundamentadas en el carácter sistemático de los lenguajes de especialidad. Esos primeros proyectos estuvieron asociados con el uso de los ordenadores personales y con las técnicas fomentadas por las bases de datos que llevaron al uso de los lenguajes coordinados en Documentación. La coincidencia mayor se da a través de las operaciones simbólicas de naturaleza lógico-semántica-pragmática, que rigen tanto la elaboración de las terminologías como la de los lenguajes documentales. Considerando en Terminología la unidad terminológica como unidad del conocimiento, y en Documentación cada elemento del lenguaje documental como

---

<sup>9</sup> Bush, V.- *As we may think*, en *Atlantic Monthly*, 1945, 176: 101-108.

<sup>10</sup> Bush, V.- *MEMEX revisited*, en *Evolution of an information society*. London: ASLIB, 1987: 182.

<sup>11</sup> Robert, N.- The pre-history of the information retrieval thesaurus, en *Journal of Documentation*, 1984, 40, nº 4: 273. Véase también Smith, L.- Memex as an image of potentiality in information retrieval research and development, en Oddy, R. et al. (ed.).- *Information retrieval research*. London: Butterworths, 1981: 345-369. Buckland, M.- Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex, en *JASIS*, 1992, 3, nº 4: 284-294.

unidades de información, se explica el funcionamiento de los sistemas de recuperación de la información que encuentran sus raíces en la asociación de ambas unidades.

El reto consistió en establecer los términos como unidades de conocimiento en sistemas de información especializados para favorecer la operatividad de los descriptores. Los tesauros actuales heredan los planteamientos determinados por la aparición en 1852 del *Thesaurus of English Words and Phrases* de Roget que, respecto a los diccionarios comunes, contextualizaba el sentido en que cada término debía ser comprendido. Mientras aquéllos van de la palabra al concepto, mostrando, por tanto, los signos que componen el sistema lingüístico de forma descriptiva, el tesoro representa de forma normalizada los conceptos de un dominio especializado mediante términos unívocos que se estructuran siguiendo principios lógico-semánticos. De manera que, si un diccionario responde cuando se le pregunta qué significa un término, un tesoro sitúa a ese término en el contexto que le da sentido, mediante las relaciones que guarda con otros términos, por lo que se establece un conjunto de expresiones para un concepto concreto, mostrándose así la proximidad conceptual que hay entre término y descriptor, pues la categorización del dominio de referencia se hace de acuerdo con las normas terminológicas, mientras que la selección de descriptores resulta de las normas documentales.

La creciente especialización del conocimiento, hacia la mitad del siglo XX, empezó a impulsar la idea del concepto como núcleo organizador de los lenguajes documentales. No se trataba ya de dar un orden físico a los impresos, si no de guiar hacia los conceptos o los significados de las representaciones de información. Desde luego, al principio de los años sesenta se impuso la idea de descriptor asociada a la de concepto (Mooers, 1963). En cuanto unidad constitutiva de un tesoro, el descriptor se define como forma de alta significación que representa a un concepto o idea y, en consecuencia, el concepto y sus relaciones se determinaron como el elemento clave de los lenguajes documentales. Precisamente en la noción del concepto coinciden un thesaurus lingüístico y uno documental.

Fue, por tanto, durante la década de los sesenta, ante la aparición de las bases de datos, cuando el tesoro pasó a tener gran protagonismo en la recuperación de la información. Se requerían soluciones que los sistemas tradicionales eran incapaces de dar, pues ahora había que representar los conceptos contenidos en los documentos, y las relaciones existentes entre ellos, en una forma estandarizada de lenguaje, con control de sinónimos y con una estructura sintáctica más simplificada que la del lenguaje natural, pero con una estructura lógico-semántica que garantizase, por inferencia, la representación del contenido informativo de los documentos de forma normalizada. Por más que el tesoro documental sea un vocabulario destinado exclusivamente a la indización y recuperación de la información, y no al conocimiento de un área, ya que la matriz de su vocabulario se basa en la reducción del número de términos a los preferentes para emplearlos en el sistema de información.

La estructura del tesoro surge de la categorización de la materia que aborda desde la que se fija su superestructura en tantos niveles como sean necesarios. En este sentido se manejan unas clases generales autoexcluyentes, condición necesaria para la organización de los términos y para el establecimiento de estrategias de búsqueda deductivas. Las categorías designan aspectos particulares de una determinada área del conocimiento, permitiendo el agrupamiento de términos, lo que les confiere autonomía de representación de un dominio dado, necesaria para organizar la información y para aspirar a unas reglas de recuperación. De manera que el tesoro debe partir de una categorización del dominio y, por lo tanto, de una

taxonomía del conocimiento temático que, en este sentido, es una terminología. Las taxonomías están presentes en todos los *Esquemas*, *Tesauros*, *Modelos conceptuales* y *Ontologías*. Taxonomía, es decir:

- Clasificación o categorización de un conjunto de descriptores de forma jerárquica, que establece una relación entre los objetos de generalización-especialización
- Una taxonomía establece una semántica simple de acuerdo con algunas de sus propiedades (Daconta; Obrst y Smith, 2003).

El procedimiento inductivo en los tesauros se cumple mediante las relaciones semánticas por las que se establecen asociaciones entre los términos que componen el corpus, permitiendo a los usuarios navegar coherentemente por los conceptos que representan:

- Integrando términos dentro de una misma categoría.
- Determinando diferencias de una especie con otras.
- Posibilitando la división o análisis del género por las diferencias.
- Concediendo definiciones para ajustar el significado de algún término.

La funcionalidad de los tesauros les ha llevado a un empleo muy exitoso, pero la ambigüedad, riqueza y capacidad de innovación constante de los lenguajes en los que se encuentran expresados los documentos sobre los que trabajan produce en ocasiones una falta de pertinencia a la hora de seleccionar documentos. A lo que se añade el problema del número de documentos que circulan por la Web y la variedad de sus soportes. De manera que su empleo ofrece indudables ventajas, pues son de creación y gestión sencillas, la eficacia ha sido muy probada, muestran además una notable coherencia, y, finalmente, son un buen punto de partida para crear ontologías, por lo que no es de extrañar que su número sea tan abundante. Aunque su uso no está libre de inconvenientes como que estén pensados para dominios restringidos, por lo que, hasta la aparición de *SKOS*, no han contado con mecanismos para compartir información en la Web; que en las jerarquías incluyan instancias, atributos y meronimias. Las relaciones son un cajón de sastre; la difícil adición de relaciones; además, se trata de un vocabulario, sin mecanismos directos que lo relacionen con los objetos documentales; no se atienen a axiomas, ni a reglas de coherencia y validación; se elaboran por consenso previo de los creadores, no de los usuarios; finalmente, el coste elevado en su creación y mantenimiento, evitable sólo con la automatización o semiautomatización.

#### **4 Presencia de la terminología en las tendencias de los lenguajes documentales**

Desde los años noventa, los presupuestos de la posmodernidad vienen obligando a que los tesauros se activen con nuevas relaciones, en número mayor e identificadas, buscando mejorar la representación y aumentar la precisión y eficacia. Han aparecido nuevas propuestas de navegación y de visualización mediante grafos explícitos de conexiones informativas. Podemos afirmar que Internet y su oferta de enlace hipertextual de documentos obligó a diferenciar la representación de los contenidos documentales, pasando a buscarse soluciones a través, primero, de los Tesauros automáticos o con la intervención de Tesauros verbales, luego con la búsqueda de mapas de conceptos que desembocó en la metodología mixta de los Topic maps.

La denominación de tesoro conceptual se fundamenta en la noción de materia (concepto) de la que tratan los textos, que aúna términos y conceptos por similitud de su sentido desde el contexto del usuario, y que se distingue precisamente por su riqueza en *relaciones asociativas* (Maniez, 1993). Puede verse asimismo como una *Red semántica conceptual* por la que se navega desde los términos más genéricos de una faceta hacia los más específicos, e inversamente (navegación vertical), que a su vez permite la transición de una clase hacia otra y de un campo de la ciencia hacia otro mediante las relaciones asociativas (navegación horizontal por nudos polijerárquicos) (Deweze, 1997). E incluso como un *Espacio conceptual* donde el tesoro aparece como un sistema formal definido por un dominio algebraico. Su modelo *espacial* define las relaciones entre términos con mayor precisión que los tesauros tradicionales. Son, pues, tesauros de carácter automático, que representan redes semánticas en la que cada nodo contiene un único concepto semántico que puede tener una serie de descriptores asociados, que pueden ser identificados según las típicas relaciones de los tesauros: preferenciales, jerárquicas o asociativas. En ellos se aprecia un incremento pragmático de las relaciones, en especial las de asociación, donde ahora se reflejan también (Tudhope et al., 2001): ideas combinadas; términos relacionados conceptualmente; contigüidad; relaciones asociativas por definición; relaciones asociativas trasladadas por significado; cuestiones de finalidad; e incluso otras sin especificar.

Respecto a los tesauros convencionales presentan estas características novedosas (Schäuble, 1987):

- a) Listan las palabras “no vacías” existentes en las bases de datos;
- b) consideran los términos coloquiales, incluso las variaciones y truncamientos de los términos reconocidos;
- c) aportan notas definitorias que aclaren las posibles dudas de uso;
- d) razonan las equivalencias existentes entre términos;
- e) contienen numerosas relaciones asociativas, incluso con los no descriptores.

En su funcionamiento, los tesauros conceptuales usan ontologías. Para mejorar la precisión de las recuperaciones documentales aprovechan el diseño de ontologías desde las que se autogeneran tesauros conceptuales que permitan distinguir los sinónimos, suprimir los homónimos e inducir relaciones asociativas entre los descriptores. Una ontología para una base de conocimientos debe abarcar los diferentes tipos de documentos, las descripciones conceptuales, las relaciones entre dichos documentos (citas), y las de éstos con los diferentes problemas científicos; además de índices, descripciones bibliográficas, tesauros, códigos clasificatorios, e información terminológica. Su aplicación debe proporcionar una visión general de la estructura y de la terminología del dominio que facilite recuperaciones relevantes.

Una ontología es una descripción formal de los conceptos y de las relaciones que existen entre los conceptos (Gruber, 1993): se componen de términos y de relaciones dentro de un dominio, junto a reglas para combinar, y de términos y relaciones para extender el vocabulario (Neches; Fikes; Finin; Gruber; Patil; Senator y Swartout, 1991). Interesa la definición de Studer, por insistir en el significado exacto de cada uno de los componentes: *Conceptualization*: modelo abstracto de un fenómeno de la realidad con sus conceptos relevantes; *Explicit*: los conceptos, sus tipos y *constraints* se definen explícitamente; *Formal*: legible por una máquina; y *Shared*: con conocimiento consensuado (aceptado por la comunidad) (Studer; Benjamins y Fensel, 1998). Una Ontología es una especificación explícita de una (o una parte de una) conceptualización, que incluye un vocabulario de términos y la especificación de su sentido.

Por tanto, una ontología especifica una forma concreta de ver el mundo, lo que se refleja en las definiciones que a su vez dependerán del lenguaje empleado en su descripción.

En los tipos de ontología siempre aparece la presencia del vocabulario en la representación. Así, en las *KR Ontologías*, los elementos en los que se expresa el conocimiento: nombres de clases, relaciones y funciones; en las *Generales (Common Ontologies)* se representan las experiencias comunes con vocabulario de cosas, sucesos, tiempo, espacio; en las *Top Level Ontologies*, mediante conceptos generales para anclar los términos raíz de otras ontologías; las *Upper-Level Ontologie*, cuentan con vocabulario y relaciones; las *De tareas* con el vocabulario necesario para realizar una tarea: verbos, adjetivos, nombres; mientras que en las de *dominios* el vocabulario es reutilizable en un dominio pero no en otros.

En los tesauros conceptuales los términos se extraen de documentos a texto completo, para luego conformar bases de conocimiento. El corpus terminológico se establece como una red semántica neuronal: en cada nodo hay un concepto semántico con el que se asocian una serie de descriptores. Los enlaces entre los descriptores pueden también establecerse según las típicas relaciones de equivalencia, facetadas o por asociación. El funcionamiento de la recuperación se establece desde la pregunta del usuario. Los conceptos que este ha solicitado se confrontan con la red terminológica, cuyos elementos están diseñados como mapas representativos de los textos, y que actúan así como lenguajes controlados que organizan la información de cualquier objeto disponible en la red. De todas formas, se observan también algunas limitaciones:

- *Desorientación de los usuarios* para encontrar los conceptos apropiados durante su navegación por los textos, debido a la saturación de enlaces en los nodos que conjuntan información de numerosas fuentes.
- Deficiente normalización terminológica de los documentos, superable mediante bases de corpus terminológicos de cada una de las lenguas más usadas.
- Las *interfaces hombre-máquina deben ser capaces de relacionar los conceptos a partir de ontologías preestablecidas*. El acceso se allanaría con representaciones gráficas de la red (tabla de contenidos gráficos con las conexiones visibles mediante enlaces hipertextuales: mapas).

Otra de las propuestas de mejora de los tesauros es la inclusión de verbos que complementen a los estáticos tesauros tradicionales de sustantivos (Levin, 1993). El uso de descriptores verbales aporta muchas ventajas, como la posibilidad de indizar audiovisuales mediante gerundios, la identificación verbal de asociaciones funcionales mucho más adaptables a dominios concretos, la posibilidad de mostrar la relación existente entre dos conceptos usando los inmensos medios del lenguaje natural (categorías verbales a modo de relaciones facetables), y la desambiguación conceptual.

La tipología de relaciones de asociación del lenguaje UML se integra así en la tendencia a ampliar el número de relaciones de asociación de los tesauros para que no planteen ninguna ambigüedad, como la relación de agregación, en la que la desaparición del todo no implica la desaparición de las partes; De composición, en la que la desaparición del todo implica la desaparición de sus partes; Información de multiplicidad (es decir, cuántos objetos pueden interactuar en una misma relación); Dirección de la relación; y Tipificación de relaciones.

Esta aproximación mediante la integración verbal procede del área pedagógica, donde esta forma de relacionar los conceptos mediante verbos recibió la denominación de mapas conceptuales (*concept maps*).

Podemos considerar como un antecedente a la base de datos de referencia léxica *WordNet*, red semántica multidisciplinar en inglés cuyo uso ha estado vinculado a la desambiguación conceptual, en especial mediante el empleo de verbos (Moldovan, 2001) y que ofrece relaciones de equivalencia y de jerarquía entre diferentes categorías gramaticales (Green, 1995):

- Se seleccionan de la frase todas las flexiones sustantivo-verbo.
- Se escoge el significado más probable del término (desambiguación terminológica):
  - Las palabras de la frase se agrupan en parejas.
  - Se buscan en WordNet los distintos significados de cada término.
  - Se forman todos los posibles pares de conceptos.
  - Se busca cada par en Internet. Luego se ordenan los resultados según los conceptos más frecuentes.
- Teniendo en cuenta los conceptos más frecuentes, se seleccionan todos los sustantivos de los “glosarios” de cada verbo y sus descendientes jerárquicos.
- Se calcula mediante una fórmula los conceptos comunes entre los sustantivos del punto anterior.
- Se ordenan todas las parejas de conceptos de sustantivo-verbo según el resultado de la fórmula.

Es directa la aplicación a la automatización en la construcción de tesauros por la aparición de sustantivos en proximidad a estructuras verbales, siguiendo un proceso que comienza con el análisis de documentos relevantes para la extracción de su vocabulario (glosarios, diccionarios, estándares sobre vocabulario, etc), luego se procede a una depuración manual del vocabulario extraído, obteniendo los descriptores, con los que se indizan los documentos textuales maestros (manuales, estándares, artículos, ...). En esta etapa se almacenan principalmente aquellas frases del documento en las que aparece uno o varios descriptores del tesoro en el Sintagma Nominal Sujeto, y uno o varios descriptores en el Sintagma Verbal. Posteriormente, los conceptos dinámicos se agrupan, clasifican y se asimilan a las relaciones del tesoro que se deseen identificar. Finalmente, se revisan a mano en el tesoro las relaciones obtenidas.

Otra innovación se ha debido a los Mapas conceptuales de navegación por redes semánticas, originados en la necesidad de crear índices en torno a algún concepto o materia. Un tipo de grafo es la red semántica, que representa las relaciones semánticas que se establecen en el texto y que no aplican ningún control de términos a la hora de establecer comunicación entre las personas y las máquinas (García Jiménez, 2004). Los grafos o redes conceptuales están constituidos por conceptos y relaciones de conceptos que son colecciones ordenadas de *nodos* (puntos/vértices) que representan conceptos y de *enlaces* (*arcs*: arcos/*edges*: extremos, satélites). Los mapas conceptuales ofrecen una red de relaciones más rica que los tesauros. El concepto que origina su estructura (nodo-enlace-nodo) es próximo a su equivalente en las redes hipertextuales por lo que soportan la navegación de un modo muy natural. El uso de mapas conceptuales permite desarrollar mejores mecanismos de representación y recuperación, ya que las relaciones entre los conceptos se eligen teniendo en cuenta las necesidades y expectativas de cada usuario. De ellos surgió la idea de establecer un nuevo

estándar conocido como *Topic map*: un documento, o un conjunto de documentos SGML o XML interrelacionados en un espacio multidimensional en el que las localizaciones son *Topic*<sup>12</sup>. Aparecen como elementos de un *Topic Map*: el *Topic*, con el *Topic type*; la *Association* y la *Association type*; el *Scope* (ámbito en el que una relación tiene sentido; y el *Topic occurrence* junto a la *Occurrence type*. Los *Topic Maps* presentan indudables ventajas, como haber optimizado los mapas conceptuales y fusionar vocabularios jerarquizados o no. Además, es un estándar ISO intuitivo tanto en su creación como en su interpretación. Por el contrario, como desventaja, no tienen inferencia, reglas ni axiomas.

Si comparásemos un tesoro (ISO 2788:1986) y un *Topic Map* (ISO 13250:2000) veríamos que estamos enfrentando sistemas en que el primero tiene un control de vocabulario a priori, frente al control de vocabulario *ad hoc* del segundo. Dentro de un dominio con validez y expansión limitada, o en referencia a dominios con validez y expansión potencialmente infinitas. Un tesoro contiene relaciones entre sus descriptores (de jerarquía, equivalencia y asociación), mientras que en el *Topic Map* se establecen usualmente mediante verbos, lo que las lleva a ser de tipo potencialmente infinito. Las relaciones de un tesoro no son descriptores, pero en un *Topic Map* una relación puede ser un *Topic*. Los términos de un tesoro están normalizados a priori, mientras que las variaciones ortográficas de un *Topic* son normalizadas a posteriori. Lo mismo sucede con la definición a priori de los descriptores, que en los *topics* se hace mediante sus *occurrences* y relaciones. De manera que los “nuevos” tesoros permiten muchos modos de asociar los términos: jerárquicos, asociativos, formales, conceptuales, referenciales, explicativos, etc., y que admiten diferentes formas de representar los conceptos, desde una línea estrictamente secuencial hasta una combinación de relaciones secuenciales, taxonómicas de distinto tipo y asociativas de diferente signo (Rodríguez Bravo, 2002).

## 5 Consideraciones finales

La terminología, como conjunto estructurado de términos, es el origen de la organización del conocimiento de un dominio especializado. Por su parte, en los sistemas de información, el descriptor, considerado como la unidad de información de los lenguajes documentales, es el eje sobre el que funcionan la indización y la recuperación de contenidos. Estructurada, una terminología no sólo agiliza la comunicación entre pares, como se cree usualmente, si no que también funciona como instrumento conceptual para la comprensión de la manera en que un dominio se organiza.

Desde la creencia ilimitada en la solución de carácter exclusivamente tecnológico a los problemas de recuperación de información que trajo la década de los noventa se frenó el desarrollo que habían tenido los lenguajes documentales fomentados por la aparición de las bases de datos referenciales en los años sesenta. Se pensaba que era poco rentable contar con lenguajes documentales que garantizaran el control lingüístico-documental de los fondos, desde una decisión tomada al preferir ofrecerle al usuario gran cantidad de datos, sin valorar su pertinencia ni su calidad. El usuario recibía ahora decepciones continuas en los resultados de las recuperaciones, sin duda porque los avances en la tecnología fueron superiores a los realizados en el grado de normalización y estructuración para manejar y reutilizar la información adecuadamente.

---

<sup>12</sup> ISO/IEC 13250: 2000. SGML-Topic Maps.

El entusiasmo inicial por la recuperación en lenguaje libre se fue diluyendo, conforme iban creciendo los problemas que su uso conllevaba (Lancaster, 1986). Surgió entonces una aspiración de incorporar y explorar los recursos de análisis y recuperación ofrecidos por una tecnología en progresión constante, que se simbolizó calificando a la Web de semántica, como reflejo de una aspiración a recuperar documentos por sus contenidos y no solo por sus expresiones de superficie.

## **Bibliografía citada**

- BACON, F. *Instauratio Magna. Novum Organum*. Nueva Atlántida. México: Porrúa, 1980.
- BARTHES, R. *Le plaisir du texte*. Paris: Seuil, 1973.
- BENVENISTE, E. *Problemas de lingüística general II*. Siglo XXI: Madrid, 1989.
- BONURA, L. *The art of indexing*. New York: John Wiley, 1994.
- CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antártica-Empúries, 1993.
- CUTTER, B. A. *Dewey Decimal Classification*. En: *Encyclopedia of Library and Information Science*. Vol. VII. New York: Marcel Dekker, 1972.
- CUTTER, C. *Rules for a dictionary catalog*. 4th ed. London: Chaucer House, Malet Place, 1962.
- DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. *The Semantic Web: a guide to the future of XML, Web services, and Knowledge management*. Indianapolis: Wiley Publishing, 2003.
- DEWEY, M. *Decimal classification and relative index*. 19th ed. Albany (New York): Forest Press, 1979.
- DEWEZE, A. *Informatique Documentaire*. 4 ed. Paris: Masson, 1997.
- FELBER, H.; PITCH, H. *Métodos de terminografía y principios de investigación terminológica*. Madrid: CSIC, Instituto Miguel de Cervantes, 1984.
- FOUCAULT, M. *L'Ordre du discours*. Paris: Gallimard, 1971.
- GARCÍA JIMÉNEZ, A. *Instrumentos de representación del conocimiento: tesauros versus ontologías*. *Anales de Documentación*, 2004, n. 7, p. 79-95.
- GONZALO, C.; GARCÍA YEBRA, V. (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros, 2004.
- GREEN, R. *The expression of conceptual syntagmatic relationships: a comparative survey*. *Journal of Documentation*, 1995, vol. 51, n. 4, p. 315-338.
- GRUBER, T. R. *A translation approach to portable ontologies*. *Knowledge Acquisition*, 1993, vol. 5, n. 2, p. 199.
- HARVEY, D. *The condition of postmodernity. An inquire into the origins of cultural change*. Great Britain: Cambridge University Pres, 1994.
- KRIEGER, M. G.; BOCOMY, M. J. *Introdução à terminologia: teoria e pratica*. São Paulo: Editora Contexto, 2004.

- LANCASTER, F. W. Vocabulary control for information retrieval. 2<sup>nd</sup> ed. Arlington: IRP, 1986.
- LARA, M. L. G. de. Novas relações entre terminologia e ciência da informação na perspectiva de um conceito contemporâneo da informação, en DataGramZero [recurso electrónico]. Revista de Ciência da Informação, 2006, vol. 7, n. 4. <[http://datagramazero.org.br/ago06/F I aut.htm](http://datagramazero.org.br/ago06/F_I_aut.htm)>
- LEVIN, B. English verb classes and alternations: a preliminary investigation. Chicago: The University of Chicago Press, 1993.
- MANIEZ, J. Los lenguajes documentales y de clasificación: concepción, construcción y utilización en los sistemas documentales. Madrid: Pirámide, 1993.
- MOLDOVAN, D. Question answering systems in Knowledge Management. IEEE Intelligent Systems, 2001, vol. 16, n. 6, p. 90-92.
- MOOERS, C. The indexing language of an information retrieval system. En: SIMONTON, W. (ed.). Information Retrieval today. Minneapolis: Minneapolis University, 1963. p. 21-39.
- MOREIRO GONZÁLEZ, J. A. Aplicación de las ciencias del texto al resumen documental. Madrid: Universidad Carlos III de Madrid; BOE, 1993.
- NECHES, R. [et al.]. Enabling technology for knowledge sharing. AI Magazine, 1991, vol. 12, n. 3, p. 36-56.
- PÉREZ HERNÁNDEZ, C. M. Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento [recurso electrónico]. <<http://elies.rediris.es/elies18/>>. [Consultado: 22 dic. 2006]
- RODRÍGUEZ BRAVO, B. El documento: entre la tradición y la renovación. Gijón: Trea, 2002.
- SAGER, J. C. Curso práctico sobre el procesamiento en terminología. Madrid: Fundación Germán Sánchez Ruipérez, 1993.
- SCHÄUBLE, P. Thesaurus based concept spaces. En: Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (20<sup>th</sup>: New Orleans, Louisiana: 1987): proceedings, p. 254-262.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: principles and methods. Data & Knowledge Engineering, 1998, vol. 25, p. 161-197.
- TEMMERMAN, R. Sociocognitive terminology theory. En: Simposio Internacional de Verano de Terminología (2º: Barcelona: 1999). Terminología y Cognición: actas. Barcelona: IULA-UPF, PPU, 2001, p. 75-92.
- TUDHOPE, D. [et al.]. Augmenting thesaurus relationships: possibilities for retrieval [recurso electrónico]. Journal of Digital Information, 2001, vol. 1, n. 8. <<http://jodi.tamu.edu/Articles/v01/i08/Tudhope/>> [Consultado: 28 dic. 2006]
- WITTGENSTEIN, L. Tractatus logico-philosophicus. Madrid: Alianza Editorial, 1982.
- WÜSTER, E. Introducción a la teoría general de la terminología y a la lexicografía terminológica. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 1998.