

Categorización de los conceptos en el análisis de contenido: su señalamiento desde la Retórica clásica hasta los *Topic Maps*

José A. Moreiro González
Jorge Morato Lara
Sonia Sánchez Cuadrado
Beatriz Ainhize Rodríguez Barquín *

Artículo recibido:
1 de febrero de 2005.
Artículo aceptado:
8 de abril de 2005.

RESUMEN

Análisis de las teorías provenientes de la retórica clásica y de la lingüística del texto sobre las que se fijan los procedimientos a seguir durante el proceso deconstructivo-reconstructivo del análisis de contenido de los documentos. Se plantea la selección de las ideas principales de los discursos siguiendo las reglas generales del cuestionar, proceso hermenéutico al que debe estar sometido todo hecho comunicativo y que tuvo en Ramón Llull las propuestas más originales. Asimismo se defiende la importancia que tiene la organización superestructural para la lectura y representación documental, y se proponen varios niveles y pasos a seguir para hacer una conveniente descripción del contenido macroestructural. La actuación del analista sobre los documentos debe hacerse de acuerdo con la propuesta de Llull, en la actualidad enmarcada bajo el paradigma

* Los cuatro autores pertenecen a la Universidad Carlos III de Madrid, España: (José: joseantonio.moreiro@uc3m.es); (Jorge: jorge@ie.inf.uc3m.es); (Sonia: ssanche@ie.inf.uc3m.es); (Beatriz: 100052527@alumnos.uc3m.es).

de Laswell (*wh questions*) y recogida por las normas internacionales que marcan el proceso de análisis. Se considera también la influencia que estas teorías han tenido sobre la estructura de los tesauros y la filosofía de actuación de los *Topic Maps*.

Palabras clave: Categorías retóricas; Representación documental; Lectura documental; Predicamentos; Aristóteles; Ramón Llull; Árbol de Porfirio; Organización del texto; Macroestructura; Superestructura; Hexámetro universal hermenéutico; Paradigma de Laswell; Teoría de Anderson.

ABSTRACT

Categorization of concepts in content analysis: its designation from Classical Rhetoric to Topic Maps

José A. Moreiro-González; Jorge Morato-Lara; Sonia Sánchez-Cuadrado; and Beatriz Ainbize Rodríguez-Barquín.

An analysis of theories based on Classical Rhetoric and Text Linguistics is made in order to set the procedures to follow during the deconstructive-reconstructive process of document content analysis. Selection of main ideas in discourse is outlined following the general rules of questioning: a hermeneutical process to which all communicational act must be submitted, and which had in Ramon Llull one of its most original proponents. The importance of superstructural organization for reading and document representation is argued, and the steps and levels to follow for a convenient description of macrostructural content are also proposed. The analyst's performance on documents must take into account Llul's proposal, today framed within Laswell's paradigm (*wh questions*) and accepted by international norms which mark the process of analysis. Also taken into account is the influence these theories have had on thesaurus structure and on the philosophy of performance of the *Topic Maps*.

Keywords: Rhetorical categories; Documentary representation; Documentary reading, Predicaments; Aristotle; Ramón Llull; Porfirio's tree; Text organization; Macrostructure; Superstructure; Hermeneutical universal hexameter; Laswell's paradigm; Anderson's theory.

INTRODUCCIÓN

Al momento de plantearnos el análisis semántico de los textos, el verdadero interés de nuestro trabajo es alcanzar a identificar y representar los conceptos capitales que contienen tales textos y que compondrán los resultados del análisis, tras aplicarles las técnicas de indización y de resumen. Los documentos de los que se ocupan los profesionales de la información son discursos que están compuestos por secuencias de mensajes a las que les da coherencia una macroestructura global. Esas secuencias pueden estar formadas por agrupaciones textuales, así como por sucesiones de otros rasgos significantes: imágenes, sonidos e incluso combinaciones multimedia. Nuestra aproximación parte de la certeza de que la comunicación documental se establece por conjuntos completos de significantes que forman discursos coherentes: los textos, las películas, etcétera, y no por partes componentes de los discursos, como las palabras o las frases. De manera que consideramos el texto como la unidad comunicativa-documental, pues solo él ofrece una secuencia completa y coherente de informaciones en el caso de los documentos escritos.

VALIDEZ DE LAS CATEGORÍAS RETÓRICAS PARA LA
REPRESENTACIÓN DE LOS CONTENIDOS DOCUMENTALES

Lo relevante de nuestra tarea es encontrar aquellos conceptos centrales en torno a los cuales obtienen coherencia los discursos, por lo tanto deberemos identificar aquella información principal cuya relevancia merezca ser representada en las entradas de un índice o en las estructuras de un resumen. En este sentido, según Aristóteles, había que considerar a los predicamentos como tipos de enunciados o formas fundamentales a las que se atienen los distintos modos de decir. El pensamiento se refiere a ellos como conceptos más generales:¹

Substancia	Cantidad	Cualidad	Relación	Lugar	Tiempo	Situación	Posesión	Acción	Pasión
------------	----------	----------	----------	-------	--------	-----------	----------	--------	--------

Esta relación fue luego reducida por Leibniz a:²

Substancia	Cantidad	Cualidad	Relación	Acción o Pasión
------------	----------	----------	----------	-----------------

1 Aristóteles, "Tratados de Lógica", en *Organon*, Madrid: Gredos, 1995. v. 1: 45.

2 G. W. Leibniz, *De Synthesi et Analyysi universali*, Berlin: Gerhardt, 1960, VII: 296.

Kant, en su *Análítica trascendental*, acabó por determinar solo cuatro categorías como conceptos puros del entendimiento:³

Cantidad:	Cualidad:	Relación:	Modalidad:
Unidad	Realidad	Sustancia y accidente	Posibilidad-imposibilidad
Pluralidad	Negación	Causalidad y dependencia	Existencia-no existencia
Totalidad	Limitación	Comunidad o reciprocidad entre agente y paciente	Necesidad-contingencia

Sin embargo, había sido Porfirio, discípulo de Plotino, quien mucho antes, aprovechando que hacía unos comentarios sobre las categorías aristotélicas, propuso disponerlas en forma de árbol,⁴ de manera que mediante esta estructura consiguió representar didácticamente la primera red semántica, pues al poner en orden las categorías aristotélicas dejó constancia gráfica de las relaciones existentes entre los conceptos,⁵ de acuerdo con su género, subtipo y diferencia:

Género Supremo:	Sustancia
Dif. genérica:	material / inmaterial
Gen. Subalterno:	Sustancia corpórea
Dif. genérica:	animada / inanimada
Gen. Subalterno:	Viviente
Dif. genérica:	sensitivo / insensible
Gen. Próximo:	Animal
Dif. específica:	racional / irracional
Especie:	Hombre
	Miguel, Manuel, Isabel, etcétera.

Podemos observar que entre las ideas universales existe un orden jerárquico. El género supremo de los universales es la sustancia material o compuesta, y desciende en la escala jerárquica de los universales de acuerdo con el orden marcado en el árbol para las categorías aristotélicas por *Genus* y *Species*. Este orden ha pervivido hasta alcanzar la estructura conceptual de los tesauros, que mantienen esa relación categórica entre los términos mediante

3 J. Bennett, *La "Crítica de la razón pura" de Kant*, Madrid: Alianza Editorial, 1979, v. 1: *La Analítica*, Versión del inglés de A. Montesinos.

4 J. Ferrater Mora, *Diccionario de filosofía*, Nueva edición revisada por Joseph María Terrabras; supervisión de Priscilla Cohn Ferrater, Barcelona: Ariel, 1999, v. 1: 49.

5 J. F. Sowa, *Knowledge representation: Logical, Philosophical and Computational Foundations*, Pacific Grove: Brooks / Cole Thompson Learning, 2000.

J. A. Moreiro, *et al.*, "Mapas conceptuales, *topic maps* y tesauros", en *JOTRI*, febrero 2004, <http://congrega.fund.uc3m.es/jotri2003/ponencias/mapas.pdf> (Consultado 08.04.05).

Genus (*Top Term* o Macrodescriptor en un tesaurus) y *Species* (Específicos de diferente nivel en un tesaurus).

- Género supremo: *Top Term*.
- Géneros y especies subordinados (intermediarios): *Middle Term*.
- Especie especialísima: Genéricos.
- Individuos: Específicos.

Cada género inferior tiene al género inmediato superior como genérico, del cual es especie, pero a su vez es genérico de los géneros que le siguen. De manera que una misma idea puede ser género (por su relación con las ideas a las que se extiende su concepto) y especie (por su relación con otra más general) hasta alcanzar a las ideas más comunes.

La categorización hecha por Porfirio, más que la original de Aristóteles, fue aceptada por Ramón Llull, quien además situó como metodología fundamental *la quaestio* o pregunta en el centro de su *Logica nova*. El árbol que propuso Llull constaba de siete partes de las que nos interesan tres: la estructura del árbol, los predicables y las categorías o predicamentos. Precisamente sobre la primera distinción, compuesta por *ens, substantia, cors, animal* y *quaestio*, y dentro este último apartado, Llull formuló como mecanismo universal del razonamiento las diez reglas generales de cuestionar; decámetro hermenéutico al que debe estar sometido todo hecho comunicativo:⁶

<i>Utrum</i>	→	si es o no
<i>Quid</i>	→	qué es
<i>Ubi</i>	→	dónde es
<i>Cum quo</i>	→	con quién es
<i>Quare</i>	→	por qué es
<i>Quomodo</i>	→	de qué manera es
<i>Quando</i>	→	cuándo es
<i>De quo</i>	→	de qué es
<i>Quantum</i>	→	cuánto es
<i>Quale</i>	→	cuál es

La propuesta de Llull fue recogida por Leibniz y Descartes, desde quienes alcanzó las propuestas de la Lógica de Port-Royal, movimiento que consideró que los tópicos lógicos se componían, junto a los predicables, por la Definición y la División. La definición explicaba la *quidditas* de la cosa, que

6 R. Llull, *Lògica nova: doctrina e manera d'aplicar lògica nova a la ciència de dret e de medicina*, A cura d'Antoni Bonner. Palma: Patronat Ramon Llull, 1998: 14.

se clarificaba por la diferencia de una especie con otras, la división o análisis del género por las diferencias.⁷ Volviéndonos a encontrar aquí con un reflejo inmediato en las definiciones (*Scope note*) que precisan el significado de algunos descriptores en los tesauros.

**LAS MACROESTRUCTURAS:
IDEAS PRINCIPALES DE LOS TEXTOS Y SU ORGANIZACIÓN**

Para ordenar y aglutinar los significados de las diferentes partes que conforman los textos se necesita de un significado global que disponga en torno a él todas las ideas que cada texto contiene.⁸ En el edificio textual, los elementos de base (palabras y oraciones) se agrupan bajo otros elementos de categoría superior, y se establece una subordinación de niveles semánticos por la que se ordenan sus partes. Esta organización jerárquica del texto corresponde a las macroestructuras, que representan la información contenida en un discurso (macroestructura general o global) o parte de él (macroestructuras parciales). La macroestructura general es la estructura de significado más profunda de un texto, pues representa su contenido más global: el tema y las materias de las que trata, que vinculan todos los demás elementos semánticos de dicho texto.

Estructura superficial	Microestructuras Semántica y sintaxis de las oraciones y párrafos	
Estructuras intermedias	Macroestructuras Estructuras semánticas parciales	
	Superestructura Orden de las macroestructuras en el texto	
Estructura global	Macroestructura general estructura semántica global, de dimensión ‡	<i>Sintáctica</i> <i>Semántica</i> <i>Pragmática</i>

Figura 1. Unidades de significación de los discursos

El nivel de significación más profundo lo ocupa la macroestructura de significado general o global, que representa el significado esencial del texto y

7 A. Arnaud y P. Nicole, *La lógica o el arte de pensar*, Madrid: Alfaguara, 1987: 328. De forma que, en cuanto a los predicables, la lógica de Port-Royal coincidía con Porfirio y Llull en considerar cinco predicables y no cuatro, porque incluían también la especie entre las ideas universales: géneros, especies, diferencias, propiedades y accidentes.

8 Dentro de las ciencias del discurso en el mundo Antiguo y Medieval, la Poética y la Retórica aportaron bases teóricas a la organización sintáctico-estructural de los textos, la Retórica a los textos no literarios, mientras la Poética lo hizo a los textos literarios.

le da sentido al resto de macroestructuras y microestructuras, por lo que podemos afirmar que la macroestructura global organiza el texto, ya que vincula todos los conceptos contenidos en las frases y oraciones que lo componen, actuando a modo de *superconcepto* para el resto de los conceptos.⁹ Nuestro objetivo debe dirigirse, en primer lugar, a identificar la idea de valor semántico general, la macroestructura global, para desde ella proceder al análisis del contenido más relevante de un texto, ya que las propiedades de la macroestructura global, como concepto superclase, deben seguirse en las subclases de ella derivadas, las macroestructuras parciales (y de éstas las secundarias, y luego las microestructuras), entendiéndolas como nuevos productos de un producto preexistente como idea nuclear.

Si la macroestructura global (Ma) representa el significado general de los conceptos de un texto, las macroestructuras parciales en las que cada texto distribuye jerárquicamente sus contenidos actúan sólo como submacroestructuras o macroestructuras secundarias (Ms), de las que a su vez dependen otras macroestructuras aún más parciales (mp), hasta alcanzar el nivel de significación más local, las microestructuras (mi):

$$\begin{aligned}
 Ma &= Ms^1, Ms^2, \dots, Ms^n. \\
 Ms^1 &= mp^{1.1}, mp^{1.2}, \dots, mp^{1.n}. \\
 &\quad mp^{1.1} = mi^{1.1.1}, mi^{1.1.2}, \dots, mi^{1.1.n}. \\
 Ms^2 &= mp^{2.1}, mp^{2.2}, \dots, mp^{2.n}. \\
 &\quad mp^{2.1} = mi^{2.1.1}, mi^{2.1.2}, \dots, mi^{2.1.n}. \\
 &\dots \\
 &Ms^n \dots
 \end{aligned}$$

Las macroestructuras parciales vinculan las estructuras de significación intermedia, los subconjuntos de significado diferenciado de cada parte entre las que se divide el texto; por lo que actúan como macroestructuras de esas partes y, por ello, secundarias, e incluso como macroestructuras de significado más parcial o nodular (mp). La secuencia de sus estructuras componentes tienden a formalizarse, siendo las más comunes los capítulos y párrafos en los libros, o las secuencias en las películas: graduación de presencia típica y explícita en los diferentes tipos de documentos, que resulta fundamental para su interpretación y análisis.

9 Sin la macroestructura general o global, “la coherencia de un texto sería solamente superficial y lineal”, Dijk, T. A. van, “Gramáticas textuais e estruturas narrativas”, en *Semiótica narrativa e textual*, São Paulo: Cultrix - EDUSP, 1977. Aplicado al análisis de los textos por Amaro, R. K., “Análise do discurso: textos básicos para a análise documentaria”, en Grupo TEMMA, *Análise documentaria*, São Paulo: FEBAB, 1989: 67.

Los distintos tipos de documentos siguen un modelo o forma dispositiva específica. No sólo cuentan necesariamente con una estructura semántica global, sino también con un plan esquemático global, la superestructura, estructura intermedia encargada de organizar formalmente los documentos en sucesión de macroestructuras parciales con función diferenciada. Su utilidad en el análisis de contenido se deriva de que la generación de muchos documentos sigue un modelo de esquema padrón planteado en determinadas categorías según las funciones que debe cumplir. Es así porque la superestructura ayuda a crear y a consultar un documento, pues marca el paso de significación desde la macroestructura general a cada una de las microestructuras de un discurso, y viceversa, esquematizándolo en una serie de categorías jerárquicamente ordenadas. Cada tipo de texto tiene una superestructura básica, aunque luego modificada por cada macroestructura global concreta: pensemos en el guión prácticamente fijo que sigue un cuento, una carta comercial, un noticiario televisivo, las noticias de prensa, o muchos documentos administrativos o científicos.¹⁰

Los subtextos de un texto se presentan separados físicamente de los demás subtextos vecinos mediante los subtítulos, los signos de puntuación, o un sangrado inicial menor. Si el texto se subdivide en su presentación, de forma especial en los trabajos científicos, las macroestructuras intermedias resultantes operan enlazando la macroestructura global con las microestructuras, pudiendo entenderse la redacción final como una suma de unidades intermedias. Desde el punto de vista del lector los párrafos son la vía por la que va obteniendo información de forma ordenada. Las unidades intermedias y superficiales del texto tienen que mostrar no contradicción con la macroestructura global.

Precisamente de la representación de los niveles macroestructurales surge el concepto de profundidad en el análisis de contenido. En el caso del resumen, sus modelos informativo e indicativo se diferencian por atender o no al significado de las macroestructuras parciales. Lo mismo sucede con la indización, que usará más o menos términos de descripción de un documento conforme se detenga sólo en las capas semánticas de aplicación global o profundice en las intermedias. Es evidente que no podemos indizar todos los conceptos que contiene un documento, pues desmereceríamos la necesidad de búsqueda del usuario entregándole microestructuras que pueden tener

10 En los documentos de carácter científico, la organización temática sigue una división canónica del texto en partes, capítulos, secciones y párrafos. La presentación formalizada del trabajo fue estudiada por Drop, hallando para los trabajos hechos mediante investigación empírica esta disposición arquetípica: 1. Planteamiento del problema. 2. Método de investigación empleado. 3. Resultados y discusión. 4. Conclusiones (en su caso recomendaciones). Drop, W., "Planificación de textos con ayuda de modelos textuales", en Bernárdez, E. (Comp.), *Lingüística del texto*, Madrid: Arco, 1987: 293-316.

una aparición casual o anecdótica en el texto, en definitiva que no son representativas del significado global del texto (macroestructura general) o de una parte significativa del mismo (macroestructuras parciales) y que causarían un efecto decepcionante similar al de las recuperaciones irrelevantes cuando buscamos mediante lenguaje libre en Internet. Más estúpido sería hacer un resumen que contuviese todas las estructuras significativas de un texto, ya que además de parafrasearlo, no estaríamos intermediando entre el autor y el lector, ni entregándole a éste una selección de los conceptos más relevantes.

Los discursos cuentan, finalmente, con estructuras de significación muy parcial, las microestructuras, cuya distribución semántica se localiza de forma tan limitada que sólo unifican una oración. Se trata, por tanto, de estructuras superficiales en las que se concreta la expresión,¹¹ y en este sentido tal como afirmábamos arriba, suelen ser de escaso interés para el analista documental, ya que al coincidir con las oraciones del texto, la parcialidad de su mensaje las hace irrelevantes para la representación de la información sustancial. Mediante la combinación discreta de elementos microestructurales o unidades simples, los autores constituyen unidades más complejas. También es cierto que mediante las microestructuras proceden los lectores y espectadores a acceder al contenido documental, pero salvo en las ocasiones en que una microestructura explícita lo sustancial del mensaje (por situarse en los párrafos introductorios a una obra o una sección suya, o en el caso de suponer una revisión del mensaje sustancial de toda una obra como sucede con las moralejas en los cuentos) no ofrecen interés para la representación del contenido documental sustancial.

NIVELES Y PASOS EN LA DESCRIPCIÓN DEL CONTENIDO MACROESTRUCTURAL

Respecto a los niveles de descripción de la estructura superficial, la primera derivación depende de la materia con la que esté hecho un documento (*“de qua”*), pues determina el modo diferente en que sea tratado y almacenado, de acuerdo con los diferentes soportes físicos. El tratamiento busca en este nivel la identificación del documento-objeto de información, por la que se facilita su descripción externa como objeto de información mediante los datos catastróficos: autor, título, lugar, editorial, fecha de edición, ...:

11 Soportan la armadura lingüística dentro de los límites de una oración, al formarse como el resultado de la *“elocutio”* retórica: cumplen la misión dentro del discurso de permitir el encadenamiento lineal de las oraciones en macroestructuras. Albadalejo Mayordomo, T., *Retórica*, Madrid: Síntesis, 1989: 117-126.

1. La *catalogación*, que confirma la existencia de un documento mediante sus señas de identificación física como objeto integrado en un fondo documental: se reconocen sus atributos físicos siguiendo los datos que proporciona su estructura formal. Por lo que podemos hablar de tratamiento físico según un determinado modelo que refleja los diferentes soportes materiales de cada tipo de documento. El conjunto de procedimientos que componen esta descripción externa se conoce como catalogación. Pero esta información sobre las características peculiares de un documento-objeto no satisface ningún afán de conocimiento, pues no penetra en las causas que buscan ofrecer a los posibles interesados un conocimiento riguroso y ágil del contenido que esos documentos encierran, su verdadera razón de ser.¹²
Mientras que el análisis del contenido informativo profundo es atendido por la indización y el resumen:
2. La *indización* tiene como fin identificar los documentos, haciéndolo sobre todo desde la aparición de las bases de datos por conjuntos de documentos, de acuerdo con significaciones coincidentes. Sus resultados se traspasan a unos términos de alta significación que son representativos de los conceptos macroestructurales de un documento. El objeto de la indización es guiar al usuario hasta la localización 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. La forma más común de recuperación en las bases de datos referenciales se establece mediante los descriptores, que se obtienen por aplicación de los lenguajes documentales a esferas concretas del conocimiento.
3. Y el *resumen*, cuyo fin es explicar las ideas sustanciales de un documento mediante su transmisión en lenguaje natural, empleando expresiones gramaticalmente completas y semánticamente ricas en información.

Si el problema fundamental 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, la cuestión

12 Sin duda, nuestro interés es el que originó en su día primero a la bibliografía y después a la documentación o ciencia de la información según Escarpit, R., *Teoría general de la información y de la comunicación*, Barcelona: Icaria, 1977: 120. La solución que permitió desde la antigüedad saber de qué trataban los documentos cuando iban creciendo en número hasta hacerse incontralables para el estudioso, consistió en reducir la información que éstos contenían a sus líneas esenciales. La actitud más habitual de acercarse al documento como objeto de características físicas, se ve sobrepajada para nuestros intereses cuando nos proponemos conocer la información que aquél nos traslada y que le otorga su razón de ser.

clave es ¿cómo pasar desde el texto completo a su representación documental? Para cumplirlo, la actuación del analista sobre los documentos (sea cual sea el tipo de contenidos y de soportes, por más que aquí nos reframamos con preferencia a los documentos textuales) puede considerarse en tres momentos:

- *Aproximación sensorial* al documento mediante la lectura, visionado, audición, etcétera, con el fin de identificar y comprender su información más destacada y la distribución de ésta. Todo texto distribuye su información de acuerdo con un plan funcional: los contenidos se presentan en un orden que ayuda a su lectura y comprensión.
- *Reducción a lo macroestructural*: la información del documento, tras ser interpretada, se limita a sus mensajes esenciales y actúa de forma selectiva: aquello que puede pasarse por alto se perderá para efectos documentales.
- *Nueva expresión* de los conceptos mediante términos o códigos de indización y redacción del resumen. Se trata en ambos casos de lograr un nuevo enunciado de la información fundamental del texto, buscando su comunicabilidad. Esta nueva expresión respetará las normativas existentes y se adecuará informativamente a alguno de los tipos de índices o resúmenes establecidos.

CONSECUENCIAS PARA LA LECTURA Y REPRESENTACIÓN DOCUMENTAL

Antes de ser leídos, los documentos hubieron de ser construidos, y para hacerlo los autores siguieron ciertas cuestiones clave que nos permiten a su vez derivarlas cuando queremos representarlos reduciéndolos a lo fundamental. Es así porque los discursos se desarrollan mediante la *ratio*, preservando la correspondencia entre el pensamiento, el lenguaje y la realidad extralingüística, por lo que la forma de aquélla adquiere un triple componente:

- el referente o realidad extralingüística (cuya estructura es la isomorfía),
- los conceptos o significados (cuya estructura es la isotopía), y
- la expresión lingüística o significante (cuya estructura es la isorreferencialidad).

La isomorfía general estudia cómo se unen estos componentes, para integrar la globalidad del discurso, en un proceso de categorización que consigue el molde formal de los documentos y que ha sido representado a través de unas preguntas cuya respuesta supone un alto grado de abstracción de la información.

Los universales, es decir los nombres, estados, procesos, etcétera, que aparecen en el texto, encuentran su intensionalización en las macroestructuras, que vienen a conformarse como una abstracción de las propiedades genéricas, que antes ha pasado por el proceso de formación de los conceptos o memoria semántica, a los que otorgan su forma lingüística las palabras que les dan nombre a los conceptos.¹³

Volvemos al esquema general del acto de comunicación lingüística, que surge de unas preguntas cuyas posibles repuestas llevarían, como ejemplares, a los predicados universales, palabras universales o conceptos puros mediante los que se representa la macroestructura global y las secundarias de un documento,¹⁴ y que en la práctica se logra a través de los descriptores en una indización y de la explicación de las ideas macroestructurales en un resumen.¹⁵

<i>Quis</i>	→	locus a persona	=	persona
<i>Quid</i>	→	locus a re	=	factum
<i>Quur (Cur)</i>	→	locus a causa	=	causa
<i>Quomodo</i>	→	locus a modo	=	modus
<i>Ubi</i>	→	locus a loco	=	locus
<i>Quando</i>	→	a tempore	=	tempus
<i>Quibus</i>	→	locus ab instrumento	=	facultas

Su versión más actual es conocida como las *wh questions*, una especie de inventario temático a partir del cual se le formulan preguntas al texto, por lo que podemos afirmar que funcionan como un molde categorial, como operadores que ligan una variable con su propiedad genérica, y que nuestro afán de análisis utiliza en sentido deconstructivo.¹⁶

Las preguntas ¿quién? (*who?*), ¿cuándo? (*when?*), ¿dónde? (*where?*), y sobre todo el ¿qué? (*what?*), y sus ¿cómo? (*how?*), ¿por qué? (*why?*), y ¿para qué? (*what for?*) determinan con sus respuestas la identificación de la materia que estemos tratando y sus principales circunstancias.¹⁷ A partir de ellas,

13 E. Coseriu., *Gramática, semántica, universales: estudios de lingüística funcional*, 2ª ed. rev. Madrid: Gredos, 1987: 68.

14 L. V. Bertalanffy., *Teoría general de sistemas: fundamentos, desarrollo, aplicaciones*, México: FCE, 1981: 102.

15 La gramática generativa volvió a darle importancia al hexámetro universal hermenéutico, los *septem loci* retóricos: Lausberg, E. H., *Elementos de retórica literaria*, Madrid: Gredos, 1975: 30-46. Cada pregunta es un tipo semántico y contiene todas las posibles respuestas “ejemplares” que Carnap llama “palabras universales” en su obra *Logical Structure of the World & Pseudoproblems in Philosophy*. Translated by Rolf A. George (*Der logische Aufbau der Welt*. Berlín: Weltkreis, 1928). Berkeley: University of California Press, 1969: 292.

16 T. Albaladejo Mayordomo, *Teoría de los mundos posibles y macroestructura narrativa: análisis de las novelas cortas de Clarín*, Alicante: Universidad, 1986.

17 En inglés se conocen como las *Wh-questions*: Culicover, P., *Syntax*, New York: Academic Press, 1982.

la tarea de deconstrucción junto con los demás datos ya recogidos, puede trazarse una primera síntesis del perfil del asunto en pocas líneas. Los interrogantes a los que responderá una indización son, pues, los que plantearía un usuario para saber qué utilidad le ofrece un documento. Como hemos dicho antes, obedecen a los criterios de redacción informativa tan repetidamente planteados en las décadas pasadas en los cursos de comunicación social y que se conocen como el *paradigma de Laswell*¹⁸ (*Loi de Quintilien*, para los tratadistas franceses, que ya eran conocidos en la antigüedad por haberlos propuesto Aristóteles en su *Retórica*¹⁹ y Cicerón en *De oratore*²⁰), y a los que asimismo atiende, con notable proximidad, la gramática de casos de Fillmore:

- ¿Qué?: determina las materias y asuntos de los que trata un documento. Responde a cuál es el acontecimiento principal, ¿qué sucede en el documento? Cabe considerar asimismo en este apartado a los objetos que sufren la acción, y a la manera como se presentan (estudio de..., cálculo de...,). Incluye las consideraciones causales y modales:
- ¿Por qué?: causa, motivos, antecedentes o contexto de un hecho.
- ¿Cómo?: modo o procedimiento para realizar algo. Precisa las circunstancias que rodean a la acción: las modalidades prácticas, así como la herramienta o medio a través del cual se ejecuta la acción.
- ¿Quién?: personas individuales o colectivas que protagonizan el contenido del documento, desarrollando, ejecutando o sufriendo una acción. Conviene separarlos del nombre del autor, por más que sea obvio el valor de éste como identificador de las obras y, por tanto, como elemento de recuperación. Pueden considerarse también como agentes otros seres vivos: animales o plantas.
- ¿Cuándo?: fija la fecha o la época (periodo) en la que se ejecuta la acción. Es el contexto cronológico o noción temporal de un contenido, distinto a la del atributo de edición, que también se puede considerar en la indización para búsquedas tipobibliográficas, por ejemplo.
- ¿Dónde?: lugar en el que se desarrolla la acción. Esta noción espacial o contexto geográfico debe concretarse incluso en los conceptos mínimos... como sucede en el análisis de las fotografías.

18 H. J. Lasswell., “Estructura y función de la comunicación en la Sociedad”, en Moragas, M. de (ed.), *Sociología de la Comunicación II. Estructura, funciones y efectos*, 4ª edic. Barcelona: Gustavo Gili, 1994.

19 Aristóteles, *Retórica*, Madrid: Gredos, 1983. Véase también: Albadalejo, T., *Retórica*, Madrid: Síntesis, 1989.

20 M. *Tvlli Ciceronis rhetorica*, Recognovit brevique adnotatione critica instruxit A.S. Wilkins. Oxonii: Typ. Clarendonianus, 1963. Contiene: I. Tres libros De oratore. II. Brvtvs orator. De optimo genere oratorvm. Partitiones oratoria. Topica.

- Siguiendo la propuesta de Fillmore, habría que considerar también los modificadores:²¹

- *Asociativo*: ¿con quién?
- *De situación*: ¿en qué situación se encuentra el agente/objeto?

Existe un curioso paralelismo entre estas categorías para el análisis de la información con las conceptualizaciones propuestas mediante las facetas universales de Ranganathan, método pensado para analizar por facetas el contenido de los documentos e identificar así los conceptos:²²

Personalidad	Materia	Acción	Lugar	Tiempo
¿Quién?	¿Qué? ¿Cómo?	¿Qué hace?	¿Dónde?	¿Cuándo?
Entidad	Materiales Propiedad Cualidad	Actividad		

La estructuración resultante de los mensajes por facetas puede esquematizarse de acuerdo con el siguiente esquema de funcionalidad:²³

Instrumento	¿qué medio?
Agente	¿quién ejecuta o sufre la acción?
Modo	¿cómo se realiza la acción?
Ubicación	tiempo, espacio, y situación de la acción
Resultado	beneficio o beneficiario de la acción
Finalidad	intención de la acción

Cuando analizamos documentos que aplican métodos experimentales, podemos fijar las siguientes cuestiones como ejemplos de los criterios que deberíamos retener:²⁴

1. ¿Trata el documento de algún objeto sometido a una acción?

21 La gramática de casos de Fillmore, Ch. J., "The Case for Case reopened", en *Syntax and Semantics 8: Grammatical Relations*, New York: Academic Press, 1977: 59-81. Fillmore lo determina en los niveles de: *Acción, Agente, Objeto, Instrumento, Modo, Causa, Lugar, y Tiempo*. Con los modificadores *Asociativo* y de *Situación*.

22 S. R. Ranganathan, *Prolegomena to Library Classification*, Bangalore: Sarada Ranganathan Endowment for Library Science, 1957.

23 Esta es la propuesta hecha por Isabel Ferin Cunha a partir de la gramática de casos desarrollada por Fillmore y Pottier, en su Estructuración de vocabulario, en Grupo TEMMA, *Análise documentária, A análise da síntese*, Brasília: IBICT, 1989: 63-87.

24 AENOR, "Documentación, Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización", UNE 50-121-91, en AENOR, *Documentación*, 2ª edición, Madrid: AENOR, 1997: 256.

2. ¿Contiene referencias a acciones, procedimientos, etcétera?
3. ¿Afecta esa acción al objeto?
4. ¿Habla el documento del agente que causa la acción?
5. ¿Describe instrumentos, técnicas o métodos especiales?
6. ¿Existen factores considerados en un medio o lugar particular?
7. ¿Se identifican variables dependientes o independientes?
8. ¿Enfoca el asunto desde una perspectiva particular normalmente no asociada a un campo?

Recogiendo el saber acumulado por los estudios retóricos, Anderson determinó los lugares más ricos para obtener la información relevante de los documentos,²⁵ y llegó a conclusiones que coinciden con las aportadas por la Norma ISO 5963-1985, que marcó así las zonas del documento que merecen una mayor atención:

1. El título del documento, pues la información esencial debe describirse en él. Por esta razón se escogieron los títulos como *corpus* terminológicos en los primeros programas de indización automática, para elaborar índices permutados tipo KWIC y KWOC (*Key Word in Context* y *Key Word out of Context*).
2. Los subtítulos: sumarios, tabla o índice de materias. Al recoger los títulos parciales se describe lo esencial de cada una de las divisiones del texto. Se explicita en este caso las macroestructuras parciales en disposición ordenada o, lo que es lo mismo, la superestructura del original.
3. El resumen de autor, pues su presencia en los originales facilita enormemente el análisis y la representación de los documentos. Este resumen sería suficiente por sí mismo como elemento de consulta tanto para reconocer los conceptos que representaríamos en la indización como para ser reutilizado a fin de confeccionar el resumen informativo.
4. La introducción, los primeros párrafos de los capítulos, secciones, y cualquier otra división del documento. Porque en ellas se suelen anticipar o razonar los contenidos subsiguientes.
5. Las conclusiones.
6. Ilustraciones, diagramas, tablas, y su leyenda o explicación.
7. Cualquier frase o palabra destacada por el autor en el texto mediante comillas, subrayado, cursivas, o negritas... Para indizar y resumir no sirven ni las generalidades, ni lo demasiado específico.

25 M. D. Anderson, *Book indexing*. Cambridge: Cambridge University Press, 1971.

CONCLUSIONES

La retórica clásica y la actual ciencia del texto nos suministran guías para acceder a lo sustancial de los mensajes, y funcionan como método para analizar y representar documentales. En sus propuestas sobre la estructura y enunciación de los documentos encontramos la solución a los elementos que debe contener una indización y un resumen acertados, pues nos conceden los indicadores de topicalidad para el análisis humano y también para el automático. Funcionan, pues, como conductores de la lectura selectiva que debe practicar un analista y, consecuentemente, para la representación conceptual de los documentos.

La influencia de las sistematizaciones conceptuales propuestas por la retórica ha llegado hasta nuestros días bien sea directa o indirectamente, por lo que se puede afirmar que la sistematización luliana del razonamiento sigue siendo tenida en cuenta en disciplinas como la inteligencia artificial,²⁶ las redes semánticas o la representación del conocimiento.²⁷ Siempre que se ha elaborado una propuesta para categorizar los conceptos del texto, ha aparecido la relación entre los conceptos como una forma fundamental del pensamiento y es aquí donde se encuentra la razón sobre el proceder inductivo de los tesauros tanto para establecer asociaciones semánticas entre los términos, como para permitirles a los usuarios navegar coherentemente por los conceptos que representan.

Los mecanismos generales del razonamiento proceden mediante el establecimiento de relaciones entre las estructuras de conceptos, que se reflejan en los elementos componentes de los lenguajes documentales cuando disponen de:

- Términos que se integran dentro de la misma categoría.
- Diferencias de una especie con otras.
- División o análisis del género por las diferencias.
- Definiciones (*Scope note*), para ajustar el significado de un término.

Llegamos a encontrar fundamentos del tesoro en propuestas como la hecha por Derrida en su teoría de la deconstrucción en torno a que el signo lingüístico se puede repetir.²⁸ Precisamente para evitar algunas de las

26 M. Boden (comp.), *Filosofía de la inteligencia artificial*, México: Fondo de Cultura Económica, 1994.

27 E. Trillas, *La inteligencia artificial*, Madrid: Debate, 1998.

28 El lector, que se enfrenta con múltiples textos y fragmentos del saber incluso en un mismo libro, cuando lee se cuestiona la unidad del libro, y entonces pasa a “deconstruir” como co-partícipe en la escritura. De esta manera, como estrategia tanto de escritura como de lectura, leemos y escribimos mediante un gesto “desdoblado”(repetido). En ese gesto doble, los injertos intertextuales, la significación siempre plural, la equívocidad, el juego de la *différance*, están señalando que toda práctica de lectura carece de fin. Derrida, Jacques, *La diseminación*, Traducción, José Martín Arancibia, Madrid: Fundamentos, 1975.

contingencias nietzscheanas, como la posibilidad de no decir nada o el peligro de la desapropiación del propio nombre, los tesauros acuerdan llamar inequívocamente a cada concepto por un término, para evitar cualquier anfibología o ambigüedad, y asegurar la comprensión de los significados con independencia de las situaciones.

Precisamente, el rendimiento de un tesoro descansa en su oferta de inferencia terminológica dentro de un listado de conceptos enlazados mediante relaciones semánticas. Cuando construimos o usamos un tesoro manejamos tres mecanismos:

- Un repertorio de términos sacados del campo cognitivo cuya representación se intenta,²⁹ y que plasmará los conceptos predicamentales de los originales a cuyo análisis deba aplicarse.
- Un orden proveniente de la realidad estudiada que se plasma en torno a una macroestructura de la que se siguen los esquemas y la segmentación de la colección de términos. Así, se determina una macroestructura global (el dominio de aplicación del tesoro), y unas submacroestructuras (representadas por los macrodescriptores que encabezan cada una de las subdivisiones del tesoro) desde las que se clasifica la colección de términos.
- Finalmente, unas microestructuras o estructuras básicas, integradas por los propios términos y las relaciones que los enlazan constituyendo redes semánticas.

La sintaxis del tesoro viene, pues, fijada por la macroestructura global y por la superestructura a ella supeditada. Desde ellas se disponen los patrones de identificación, ordenación, subordinación y asociación, y por tanto los diferentes campos y categorías de conceptos. De esta manera, la macroestructura cumple una doble misión: señala el fraccionamiento que se va a establecer en la colección de términos, además de ser un recurso esquemático visualizador del emplazamiento exacto de los términos dentro de la estructura del lenguaje.

Incluso los afanes por la visualización que persiguen los mapas conceptuales, como los recursos esquemáticos para representar los significados de los conceptos claves de los documentos, habían sido ya propuestos en el Árbol de Porfirio, primera red semántica o, si queremos, primer mapa conceptual. Los *Topic Maps* se distinguen, además, por las relaciones que establecen entre tópicos de textos distintos, funcionando en este caso como metaíndices. Esta

29 Los términos se ajustarán a unas propiedades formales que se muestran en la Norma ISO 2788-1986.

condición fue siempre asumida por los tesauros, pues en ellos se contienen los conceptos competentes para describir los contenidos a indizar (reconstrucción en lo esencial de un texto) en paralelo al proceso deconstructivo que refleja la inclusión de esos mismos conceptos en el original y que, por lo tanto, hace del tesoro una nómina de los conceptos que reflejan las materias propias de un área especializada sobre las que versan los documentos que la estudian (y que organizan la realidad a la que remiten) y a la que se refieren los índices y hasta los resúmenes que la representan (y que reorganizan esa misma realidad con finalidad recuperadora y explicativa).³⁰ De forma que los tesauros pueden verse como metaíndices en cuanto se componen con los términos utilizables a la hora de representar los contenidos en índices concretos.

La lingüística del texto, a través del paradigma de las macroestructuras, ha fundamentado las teorías, desde las que se establece la capacidad selectiva que implica cualquier análisis de contenido documental y que está centrada en la identificación de los conceptos fundamentales de los discursos con objeto de representarlos mediante las diferentes técnicas de indización y de resumen.

BIBLIOGRAFÍA

- Allen, J., *Natural Language Understanding*, 2nd edition, London: Benjamin Cummings Pub. Co., 1995.
- Bardin, L., *Análisis de contenido*, Barcelona: Akal, 1986.
- Campos, M. L. A., *Linguagem documentária: teorias que fundamentam sua elaboração*, Rio de Janeiro: Editora da Universidade Federal Fluminense, 2001.
- Chaumier, J., *Le traitement linguistique de l'information*, 3ª ed. París: Entreprise moderne d'édition, 1988.
- Coulon, D. y Kayser, D., "Informatique et langage naturel: présentation générale des méthodes d'interprétation des textes écrits", en *Technique et science informatiques*, 1986, 5, nº 2: 103-128.
- Dale, R., Moisl, H. Somers, H. (eds.), *Handbook of natural Language Processing*, New York: Marcel Dekker, 2000.

30 Wittgenstein, al analizar el lenguaje y su relación con el mundo, concluyó que el lenguaje organiza la realidad, con lo cual se da la posibilidad de adecuación e inadecuación entre los signos y los símbolos. Wittgenstein, Ludwig, *Tractatus logico-philosophicus*, Versión e introducción de Jacobo Muñoz e Isidoro Reguera, Madrid: Alianza Editorial, 2003. Wittgenstein confirmó que la tarea de la filosofía no consistía en corregir el uso ordinario del lenguaje, sino en comprender su funcionamiento de forma adecuada, acabando así de resolver, por disolución, los problemas filosóficos tradicionales. Wittgenstein, L., *Investigaciones filosóficas*, traducción de *Philosophische Untersuchungen*, por Alfonso García Suárez y Ulises Moulines, Barcelona: Altaya, 1999.

- Dijk, T. A. van, “Gramáticas textuais e estructuras narrativas”, en *Semiótica narrativa e textual*, São Paulo: Cultrix-EDUSP, 1977.
- Dijk, T. A. van, *La ciencia del texto: un enfoque interdisciplinario*, Barcelona: Paidós, 1983.
- Donlan, D., “Locating main ideas in history textbooks”, en *Journal of Reading*, 1981, 24, n° 2: 135-140.
- Eco, U., *La estructura ausente*, Barcelona: Lumen, 1978.
- Gil Urdiciain, Blanca, *Manual de lenguajes documentales*, Gijón: TREA, 2004.
- Iwanska, L. M. y Shapiro, S.C., *Natural language processing and knowledge representation: language for knowledge and knowledge for language*, Menlo Park: AAAI Press: The MIT Press, 2000.
- Lefèvre, P., *La recherche d'informations: du texte au thésaurus*, Paris: Hermes, 2000.
- Moreiro González, J. A., *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural*, Gijón: TREA, 2004.
- Moreiro González, J. A., *Aplicación de las ciencias del texto al resumen documental*, Madrid: Universidad Carlos III de Madrid-Boletín Oficial del Estado, 1993.
- Serrai, A., “Lingua naturale e linguaggi documentari”, en *Ricerche di Biblioteconomia e di Bibliografia*, 1983, 25: 55-72.
- Singer, H., y Donlan, D., *Reading and Learning from Text*, Hillsdale, New Jersey: Erlbaum, 1989.
- Talamo, M. F.; López, M.; Yumiko Kobashi, N.; Keiko, R., “La interface: análisis documental, lingüística documental y terminología”, en *Cuadernos de la Adab*, 1993, 1, n° 2: 381-388.
- Verdejo Maillo, M.F., “Comprensión del lenguaje natural: avances, aplicaciones y tendencias”, en *Procesamiento del lenguaje natural*, 1994: 5-29.

