

## Capítulo 1

### ANÁLISIS DE CONTENIDO E INTERMEDIACIÓN DOCUMENTAL

El hablar «registrado» de los documentos se especifica en muchas maneras articuladas que han creado los autores para comunicarse intencionalmente con otras personas.<sup>1</sup> Esas articulaciones de nuestro interés son los documentos, cuyos contenidos informativos corren fundamentalmente por el cauce del lenguaje. Así pues, generalizando, podemos afirmar que el hecho documental es un acto de comunicación lingüística; consiste en que un autor emite un mensaje grabado con la intención de que alcance a potenciales receptores, que lo reciben en un contexto dado, a través de un canal y mediante un código común con el autor. De esta forma, el hablar de los documentos se establece como actividad comunicativa social e intencional, que subyace y precede a cualquier actividad creativa o científica para la que los documentos se idean. Desde una perspectiva científica, este cuadro comunicativo supone que los signos no existen si no se comunican.

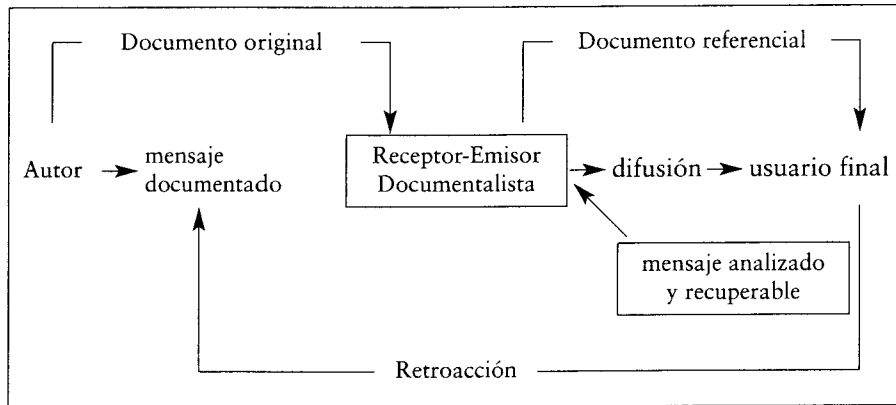
En nuestro campo de actuación (la documentación) partimos, pues, de un objeto semántico, la información, que se expresa a través del lenguaje en articulaciones concretas que llamamos *documentos*. Los modelos teóricos lingüísticos que estudian los textos nos permiten interpretar esas unidades, al tiempo que nos conceden herramientas para proceder a su análisis y representación. Contamos con mensajes completos expresados en documentos cuya comprensión y descripción no puede hacerse lejos de la lengua, de tal manera que la organización estructural, los medios analíticos y la representación del nuevo producto documental se realizan mediante el lenguaje.

<sup>1</sup> Siguiendo la definición aristotélica, Valerio Báez propone que el hablar es: comunicarse con alguien intencionalmente, de manera articulada. (*Desde el hablar a la lengua. Prolegómenos a una teoría de la sintaxis y la semántica textual y oracional*, Málaga: Ágora, 2002, p. 17).



sintáctica determinada. Mientras que la estructura semántica o abstracta atiende a comunicar la información enunciada en el documento.<sup>3</sup>

Fig. 2. Intermediación informativa de la documentación



El cuadro operativo y transformativo en el que se incluye nuestro objeto de estudio puede esquematizarse a partir de una información original o primaria, que es el documento íntegro tal como lo generó el autor y como se difunde desde las editoriales, para alcanzar una información referencial o secundaria de aquella, mediante representaciones del contenido conseguidas al hacerse el análisis documental.<sup>4</sup>

Respecto a los niveles de descripción de la estructura superficial, una primera derivación surge en dependencia de la materia *de qua* esté hecho un documento, determinando 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 catalográficos (autor, título, lugar, editorial, fecha de edición, etc.), dándose así la primera operación fundamental por traspasar la carga informativa de los textos:

<sup>3</sup> F. Saussure de: *Curso de lingüística general*, Buenos Aires: Losada, 1965.

<sup>4</sup> Hjelmslev abrió la teoría del signo lingüístico: a partir de la función de interdependencia entre el contenido y la expresión, distinguió en cada uno de estos dos componentes los aspectos de *forma* y *sustancia*. De modo que estableció cuatro estratos con dos pares correlativos: sustancia de contenido y forma del contenido; sustancia de la expresión y forma de la expresión. L. Hjelmslev: *Prolegómenos a una teoría del lenguaje*, Madrid: Gredos, 1986.

1. La catalogación o descripción externa confirma la existencia de un documento y le concede sus señas de identificación física como objeto integrado en un fondo documental. Se reconocen los atributos del documento mediante la recogida de los datos de su estructura formal. Se constituye así el tratamiento físico, desde su consignación a un determinado modelo de acuerdo con las diferentes formas materiales en que esté confeccionado o dispuesto cada tipo de documento. El conjunto de procedimientos y el resultado de la descripción externa se denomina generalmente *catalogación*. Pero esta transmisión de las señas peculiares que distinguen el documento como objeto diferenciado no satisface nuestro afán de información, pues no penetra en las causas que realmente nos preocupan y que en su día dieron origen primero a la Bibliografía y después a la Información-Documentación<sup>5</sup> y que buscaban ofrecer a los posibles interesados un conocimiento lo más riguroso y ágil del contenido que esos documentos encierran, su verdadera razón de ser.

Mientras que al análisis de la estructura informativa profunda atienden, por lo que respecta a la superestructura del original, la tipología textual (también la métrica y estilística literarias), siendo el contenido de las macroestructuras analizado por la indización y el resumen:

Fig. 3. Niveles de análisis documental

Documentos originales	Niveles de descripción		Niveles de análisis	Representaciones referenciales
	Estructura superficial	Soporte	Catalogación	
		Identificación del documento-objeto		
	Estructura semántica	Superestructura	Tipología textual	
Macroestructuras		Indización		
		Resumen		

<sup>5</sup> R. Escarpit: *Teoría general de la información y de la comunicación*, Barcelona: Icaria, 1977, p. 120. La solución que permitió desde la antigüedad saber de qué trataban los documentos, cuando estos iban creciendo en número hasta hacerse incontrolables para el estudioso, consistió en reducir la información que 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 aquel nos traslada y que le da razón de ser.

2. La indización o descripción característica, cuyo fin es la identificación de los documentos por conjuntos, de acuerdo con significaciones coincidentes. Sus resultados reflejan en términos significativos y representativos los conceptos que contiene el documento. Sobre esos descriptores del contenido, eje procesual de nuestra ciencia, se sitúan los métodos más aceptados para recuperar la información. Heredera de las técnicas de indización cuya filosofía ha sido siempre guiar al lector hasta la ubicación exacta de la información que persigue, hoy la recuperación documental es la elección de las obras existentes sobre un asunto científico o general cuya información se adecua a las necesidades del demandante. 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, necesitándose además la combinación de esos elementos a través de las operaciones de álgebra de Boole, así como de una aplicación por la que se establece el proyecto de recuperación, la estrategia de búsqueda. La adecuación entre los términos que describen fenómenos de un lenguaje especializado y la denominación de los hechos narrados por un texto concreto viene dada por un instrumento del lenguaje documental, el tesauro, primera norma de pertinencia y relevancia de los descriptores.
3. Y el resumen o descripción sustancial, cuyo producto consigue la reducción del mensaje mediante la expresión más completa, rica y explicativa del documento plenamente desarrollado. Extrae las estructuras semánticas fundamentales y las expresa conforme a los postulados del lenguaje natural. 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.

### 1.1. Fases del proceso

En el análisis de contenido hay una cuestión clave: ¿cómo pasar desde la unidad del texto completo a su representación documental? Para cumplir este proceso, la actuación del analista sobre los documentos (sea cual sea el tipo de contenidos y de soportes) puede considerarse en tres momentos:

- *Reconocimiento*: aproximación sensorial al documento mediante la lectura, visionado o audición con el fin de identificar y comprender su información más destacada y la distribución de la misma. Todo texto distribuye su información de acuerdo con un plan funcional: los contenidos se presentan en un orden (superestructura) que ayuda a su lectura y comprensión.
- *Reducción*: la información del documento, tras ser interpretada, se limita a sus mensajes esenciales. Aquello que puede pasarse por alto se perderá a efectos documentales. El procesamiento documental de los mensajes exige que se aminoren los datos de acuerdo con la importancia de su significado. Para poder manejar los documentos hay que limitarlos a sus líneas macroestructurales.
- *Representación*: expresión de los términos o códigos de indización, y redacción del resumen. Se trata en ambos casos de lograr una nueva expresión 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 índice o resumen establecidos.

### 1.2. Texto y documento

Comprendemos como documentos aquellos discursos cuya secuencia de mensajes posee una macroestructura integradora, formándose esta secuencia generalmente por agrupaciones gráficas, pero también por sucesiones coherentes de imágenes, sonidos y otros rasgos, lo que supera la hasta ahora común referencia casi exclusiva al documento escrito. De esta forma entramos obligatoriamente en una semántica documental, pues establecemos nuestros estudios sobre la información contenida en sucesiones de mensajes.

Un texto, un documento, no tiene sentido si no es informativo (si no contiene una cantidad determinada y coherente de datos) y si no es a la vez potencialmente comunicativo. No podemos hablar de *información del texto* si no consideramos cumplida la actividad comunicativa del mismo. Comunicación e información se implican indisolublemente. La actividad comunicativa del texto traslada información de manera ordenada, que no se realiza hasta que las proposiciones emitidas por un autor no alcancen a quienes van dirigidas.<sup>6</sup>

<sup>6</sup> Precisamente la lingüística textual no puede dejar de comprender al «lenguaje como comunicación». E. Bernárdez: *Introducción a la lingüística del texto*, Madrid: Espasa-Calpe, 1982, p. 57.

Nos comunicamos no por palabras o frases, sino por textos. El texto queda así considerado como unidad comunicativo-documental, en cuanto secuencia completa y coherente de informaciones. Sin función comunicativa el texto es absurdo, el documento no existe. Por ello no podemos comprenderlos solo con criterios lingüísticos. Cada texto se integra en un sistema más amplio que fija su contexto extralingüístico. Cada discurso científico no es sino un fragmento de un discurso general que viene argumentado por cada ámbito científico y su evolución, y por una metodología que casi siempre define la superestructura.

La dicotomía significado/significante se traduce al texto por un paralelismo en el cual las oraciones y frases que lo conforman resultan de la unión de los planos de expresión (lo sintáctico), del contenido (lo semántico) y de la relación entre comunicantes y de estos con el mensaje (lo pragmático). Ninguno de ellos puede ser olvidado a la hora de plantearse un análisis en profundidad, si bien nuestra referencia al discurso se origina en su consideración global, como conjunto de secuencias de microestructuras, y por ello debe valorarse más en tanto que se refiere a los contenidos, al carácter semántico.

### 1.3. La organización de los textos: las macroestructuras

Los textos se disponen en partes que se diferencian con claridad a través tanto de señales sintácticas o léxicas, como de variaciones en los esquemas semánticos. Para ordenar y aglutinar los significados de esas diversas partes el texto necesita tener un significado básico o general<sup>7</sup> por lo que nuestro primer interés consistirá en alcanzar los valores semánticos generales para, partiendo de ellos, poder efectuar el análisis.

En la arquitectura textual, los elementos de base se unen para formar otros elementos de categoría superior, de tal manera que el texto viene a conformarse mediante una subordinación funcional, especie de niveles por los que se ordenan sus partes. Las macroestructuras son las encargadas de organizar jerárquicamente el texto. Una macroestructura es una representación de la información contenida en

<sup>7</sup> Poética y retórica han sido las ciencias clásicas del discurso. La retórica aportó un importante cuerpo teórico a la organización sintáctico-estructural de los textos no literarios en el mundo antiguo y medieval, mientras que la poética hacía lo mismo para los textos literarios. Sus aportaciones se revisaron con el movimiento neorretórico de las décadas de 1960 y 1970, cuando entraron a formar parte de los conceptos manejados por la estructura global de los textos.

Fig. 4. Unidades semánticas del discurso

Estructura superficial	Microestructuras: secuencia de oraciones concretas		
Estructuras intermedias	Macroestructuras: estructura semántica parcial		
	Superestructura: disposición de las partes del discurso		
Estructura global	Macroestructura global: estructura semántica global, de dimensión	<i>Sintáctica</i>	<ul style="list-style-type: none"> <li>• estructura temática transformacional,</li> <li>• sintaxis de base lineal.</li> </ul>
		<i>Semántica</i>	<ul style="list-style-type: none"> <li>• contenido global del discurso, su tema.</li> </ul>
		<i>Pragmática</i>	<ul style="list-style-type: none"> <li>• función comunicativa;</li> <li>• generación y comprensión textual (macrorreglas y estrategias);</li> <li>• coherencia y cohesión del discurso.</li> </ul>

un discurso o parte de él. Se corresponde con la estructura profunda, representación semántica global que define el significado del discurso, el tema y las materias de las que entiende.

En el nivel de significación más profundo, nos encontramos con una macroestructura que representa el significado global del texto. Solo desde esta macroestructura global pueden comprenderse los textos, pues ella da sentido a las proposiciones que estos contienen.<sup>8</sup> Es precisamente la visión global de un discurso la que permite, primero, organizar al autor y, luego, comprender su mensaje por los lectores. La macroestructura organiza el texto, ya que como idea más general o global vincula todas las afirmaciones hechas en las frases y oraciones que lo componen. La macroestructura contiene la significación esencial y, por ello, global del

<sup>8</sup> T. A. van Dijk: *La ciencia del texto. Un enfoque interdisciplinario*, Barcelona: Paidós, 1983, pp. 187-199.

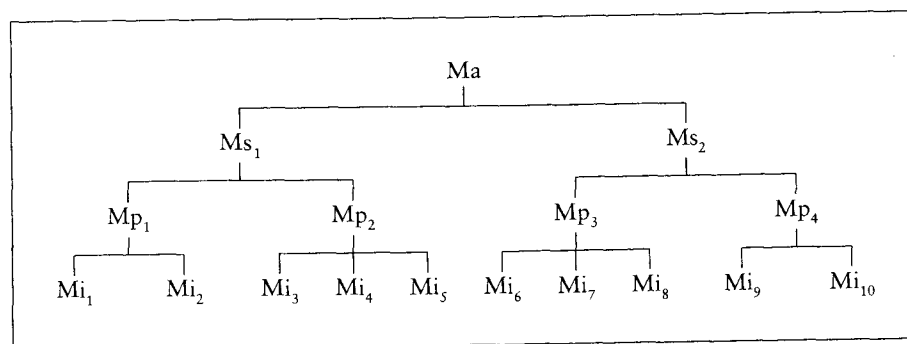
discurso. Sin esa macroestructura, «la coherencia de un texto sería solamente superficial y lineal».<sup>9</sup>

Desde nuestros propósitos, nos interesa insistir en la macroestructura general como proposición informativa que vincula las otras macroestructuras de ese texto. La macroestructura global (*Ma*) es el significado total de las diversas macroestructuras parciales en que el texto se organiza jerárquicamente, por lo podemos denominar a estas submacroestructuras o macroestructuras secundarias (*Sm*), de las que a su vez dependen otras macroestructuras parciales (*mp*), cuyo nivel más local sería el de las microestructuras (*mi*).

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

Las estructuras del texto se ordenan *jerárquicamente*, hasta alcanzar la más general, de forma que la información de una secuencia de frases se integra de modo natural en una unidad semántica de representación global.

Fig. 5. Estructura formal de la base textual



<sup>9</sup> R. K. Amaro: «Análise do discurso: textos básicos para a análise documentaria», en GRUPO TEMMA: *Análise documentaria*, San Paulo: FEBAB, 1989, p. 67. Sobre Teum A. van Dijk: «Gramáticas textuais e estruturas narrativas», en *Semiótica narrativa e textual*, São Paulo: Cultrix-EDUSP, 1977.

- Ma*: Macroestructura de significado general o global. Toda macroestructura global es un *superconcepto* para otros conceptos. Vincula todas las proposiciones de un texto o discurso.
- Ms*: Submacroestructuras, o macroestructuras secundarias. Vinculan la información de cada parte entre las que se divide el texto.
- Mp*: Macroestructura de significado nodular o parcial.
- Mi*: Estructura de superficie: secuencia de oraciones concretas de significado local.

Cuando las macroestructuras actúan como armaduras intermedias del texto, nos encontramos con las macroestructuras parciales. Cada una de ellas especifica la información del texto en subconjuntos de significado diferenciado, pero coherente. Estas unidades tienen a formalizarse como secuencias de oraciones, siendo las más comunes los capítulos y párrafos: graduación de presencia típica y explícita en los documentos científicos, y que resulta fundamental para su interpretación y análisis.

Finalmente, contamos con estructuras de significación muy específica; la distribución semántica es en ellos tan parcial que solo unifica una oración, por lo que se denominan microestructuras. Son el resultado de la *elocutio* retórica:<sup>10</sup> soportan la armadura lingüística dentro de los límites de una oración. Por lo que podemos afirmar que las microestructuras coinciden con las oraciones del texto, tanto en lo sintáctico como en lo semántico. Su misión dentro del texto es permitir el encadenamiento lineal de las oraciones en macroestructuras. A través de ellas, el autor da forma explícita a la composición que desarrolla sobre las macroestructuras, y se comprenden estas por los lectores.

#### 1.4. Superestructura e identificación de las partes del texto

La gramática generativa relanzó el concepto de *superestructura* que luego emplearon los lingüistas del texto para referirse a los modelos o formas textuales específicas de los distintos tipos de textos. Los textos no solo tienen una estructura semántica global, sino también un plan esquemático global, la superestructura,

<sup>10</sup> La microestructura es la estructura superficial en la que se concreta la expresión. T. Albadalejo Mayordomo: *Retórica*, Madrid: Síntesis, 1989, pp. 117-126.

estructura intermedia encargada de distribuir el texto en partes con función diferenciada. Su misión es por tanto organizar formalmente el texto. Su utilidad en documentación se deriva de que la generación de muchos textos se rige por ciertos modelos.

Sea dentro del texto  $f$  la función que cumplen los segmentos textuales, entonces expresáramos la superestructura como:

Texto,  $f^a/\text{texto}$ ,  $f^b/\text{texto}$ ,  $f^c/\text{texto}...$  /  $f^n$ .

El texto muestra diversas categorías según las funciones que debe cumplir. Las superestructuras ayudan a producir, leer y comprender un texto, pues lo esquematizan en una serie de categorías jerárquicamente ordenadas. Cada tipo de texto tiene una superestructura básica: pensemos en el guión prácticamente fijo que sigue una carta comercial, un cuento, un prospecto farmacéutico, las noticias de prensa o muchos documentos científicos. Sobre todo en este último caso, el receptor alcanza la organización temática a través de la división canónica del texto en unidades temáticas tales como partes, capítulos, secciones y párrafos. Cada una de estas divisiones poseerá una macroestructura parcial propia con dimensiones sintáctico-semánticas plenas, que se relaciona explícita o implícitamente con la macroestructura general. Es notorio cómo los documentos de archivo suelen atenerse a un modelo textual muy estable, lo que facilita su procesamiento automático, prácticamente desde el momento de su confección.

La superestructura de un texto narrativo presenta este esquema típico:<sup>11</sup> *historia* (planteamiento), *complicación* (trama), *desarrollo* (episodios y marco), *evaluación* (suceso-compilación-resolución) y *moraleja*. Estas categorías se ven como funciones específicas que se asignan a las distintas macroproposiciones textuales. Lo que se quiere demostrar es la moraleja (núcleo); la historia (satélite) es la justificación de la moraleja.

Mientras que la división arquetípica del trabajo científico se puede ejemplificar a través del modelo global de Drop,<sup>12</sup> que en el caso de un proyecto de investigación empírica se dispondría así:

- planteamiento del problema;
- método de investigación;

<sup>11</sup> T. A. van Dijk: *Texto y Contexto. Semántica y pragmática del discurso*, Madrid: Cátedra, 1980: 68-70.

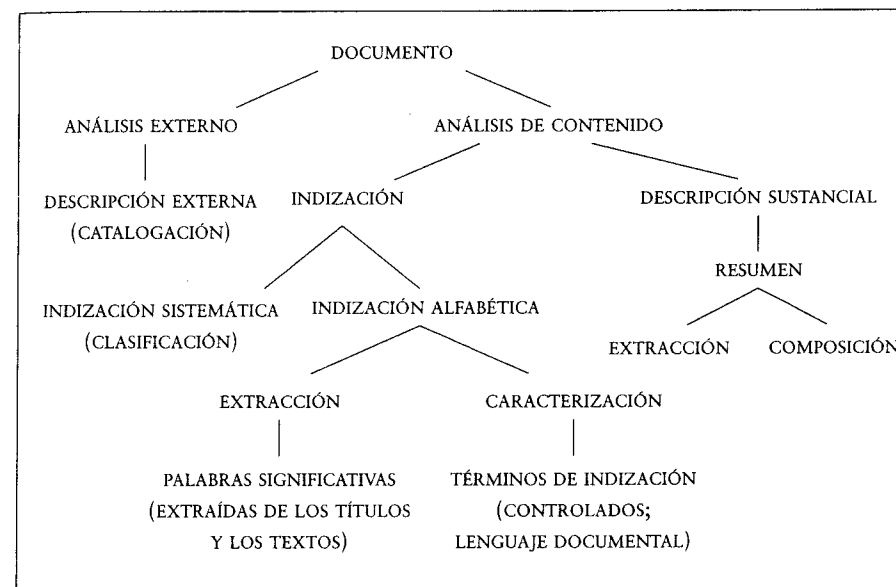
<sup>12</sup> W. Drop: «Planificación de textos con ayuda de modelos textuales», en E. Bernárdez (comp.): *Lingüística del texto*, Madrid: Arco, 1987, pp. 293-316.

- discusión y comparación de resultados;
- conclusiones (en su caso recomendaciones).

Por más que esta convencionalidad varíe dependiendo de las soluciones que se hayan de dar a los problemas que se traten, no podemos olvidar que todo discurso científico tiene carácter de acción.

Las secciones del texto suelen presentarse físicamente distinguidas de los demás párrafos o capítulos vecinos a través de los signos de puntuación, de subtitulación o de un sangrado inicial menor. En las aportaciones científicas, sobre todo, estas macroestructuras intermedias enlazan la macroestructura global con las microestructuras. El trabajo final es una adición de unidades intermedias. Desde el punto de vista del lector, los párrafos han sido considerados como el camino por el que se describe la estructura y las características de las partes. El receptor va obteniendo información de forma ordenada.

Fig. 6. Proceso de análisis de los documentos



Los esquemas imitan la forma en que se almacena el conocimiento en la memoria (términos semánticamente relacionados, sinónimos, elementos comunes a cosas...). La información relativa a un esquema se entiende sin necesidad de hacerla explícita. Suelen ser fragmentos del mundo sobre los que estamos hablando. De esta manera podemos ver el texto desde sus partes, que vienen diferenciados por los esquemas o guiones y que es el proceso que sigue la inteligencia artificial: intentar entender las partes del texto. La densidad léxica refleja la estructura semántica y nos permite identificar de qué tratan los textos y las partes de estos.

Veremos luego cómo los modelos informativo e indicativo del resumen obedecen a una profundización o no en las macroestructuras parciales, nivel que en paralelo se traslada también a la indización. Esta contendrá más o menos términos de descripción de un documento conforme descienda a estas capas semánticas intermedias o se quede solo en las de aplicación global.

## Capítulo 2

### EL RECONOCIMIENTO O LECTURA DEL DOCUMENTO

De forma ideal, la comprensión completa de los documentos impresos requiere de una consulta detallada. Sin embargo, una lectura completa es a menudo impracticable y no siempre necesaria, por más que el analista deba asegurarse de que no ha descuidado ninguna información útil. Resulta obvio que la lectura del analista difiere de la de un lector normal en varios aspectos: el analista se enfrenta a los documentos con un tiempo limitado y su lectura se orienta preferentemente a la tarea profesional de catalogarlo, caracterizarlo, resumirlo, asignarle descriptores o palabras clave, etcétera, en lugar de a la comprensión y asimilación del mensaje que pretende transmitir. Evidentemente, los estándares ISO formulan recomendaciones o métodos para el examen de los documentos, la determinación de sus materias y la selección de los términos de indización. En ellos se sugiere un acercamiento general al proceso de análisis, pero no constituyen un procedimiento riguroso que produzca resultados predecibles. Las indicaciones que se hacen en ellos son útiles, pero se limitan en exclusividad al documento, sin ofrecer explicaciones específicas sobre las diferencias entre distintas disciplinas o grupos de usuarios.<sup>1</sup>

La lectura de un texto con finalidad analítica-documental debe caracterizarse por ser, en primer lugar, de situación, y, durante todo su proceso, activa. La lectura de situación consiste en ir planteando al texto unos interrogantes antes de comenzar y conforme se avanza en la lectura. El primer paso trata de situar contextualmente la obra analizada: ¿conozco algo del autor, de su obra, de sus ten-

<sup>1</sup> B. Hjørland concluye en su trabajo: «Information retrieval, text composition, and semantics», *Knowledge Organization*, 1998, 25, núms. 1-2, pp. 16-31, que las normas ISO/BSI consisten en categorías muy genéricas al estar pensadas para su aplicación a la inmensa mayoría de los documentos, pero que no atienden las necesidades específicas asociadas a dominios particulares de la actividad humana.

dencias? También debemos identificar la naturaleza del texto que manejamos: ¿en qué tipo de documento podemos incluirle?

La lectura activa es una disposición que parte de la actitud del documentalista frente a la información a cuyo análisis se enfrenta. Desde luego, su aproximación a un texto es bien distinta de la que efectúan las demás personas cuyo afán se limita solo a la consulta. Existen unas guías mentales que ayudan a mantener la atención hacia el fin que perseguimos. Cuando consideramos esto vamos estando inmersos de forma simultánea en el paso siguiente. Parece aceptarse de modo general que el analista sigue un proceso que consta, primero, de una percepción sensorial de los documentos (lee, ve, examina, escucha...), luego interpreta su mensaje mientras lo entiende (influido por su experiencia previa y sus conocimientos que incluyen su interpretación de las instrucciones recibidas) y, finalmente, describe su versión del mensaje de acuerdo con las reglas y patrones del tipo de índice sobre el que trabaja. No puede olvidarse en nuestro ámbito la orientación que todos los procesos deben tener hacia el usuario. El trabajo del analista debe tener siempre presente la perspectiva de la audiencia a la que se dirige. El principal objetivo de esta parte del trabajo profesional es adelantarse a las demandas de los usuarios sobre la información contenida en los documentos.

La lectura y el análisis deben hacerse con mucha atención. Buscamos discernir lo fundamental de lo accesorio y graduar las unidades de contenido en que el original se estructura. Desde que leemos el título, comienza una continua búsqueda de la idea principal. Lo más normal es que la idea esencial se exprese al principio del texto. Cuando tenemos la seguridad de haberla alcanzado, la sucesión en la lectura nos irá garantizando que hemos acertado. Sintetizar la información de un documento es una muestra de habilidad. Así lo requieren los conceptos que se deben reformular, condensar o precisar. No es fácil elaborar una fórmula que permita alcanzar esa habilidad. Sin embargo, podemos encuadrarla dentro de un marco de partida mínimo.

Un texto representa la progresión de un pensamiento. Esta es la razón de que se le pueda llamar discurso. Al leer hay que poner en claro las etapas de esta progresión, que luego se reflejarán en el resumen. Para discernir lo fundamental de lo accesorio, y para graduar las unidades de contenido en que el original se estructura es preciso darse cuenta de los enlaces, tanto de los explicitados por palabras, como de los de carácter lógico entre las ideas del texto. Por lo cual es necesario fijarse en cómo se hacen depender y trabar las ideas secundarias con la idea principal. Tras identificarlas, el encadenamiento de ideas del texto se deberá respetar en la representación.

Hay que considerar como enlaces explícitos las conjunciones de coordinación y subordinación, los pronombres relativos, y los adverbios de tiempo y de modo.

### 2.1. Procesos inferenciales en la lectura

Debemos siempre tener presente que indizar o resumir es como trillar: exige separar lo sustancial de lo accesorio. Hacia esa adquisición de la información de modo selectivo nos conduce la lectura activa. Conforme vayamos leyendo, iremos subrayando las expresiones más complejas y localizando las palabras clave de un texto. Podremos así determinar los puntos que contienen la idea principal de cada párrafo. Y siguiendo el razonamiento del autor extraeremos la organización del documento. El propio texto nos llevará hacia ese objetivo, no solo mediante la identificación de palabras, sino a través de las locuciones de engarce entre las oraciones, y valiéndonos de las redes semánticas. Parece casi innecesario comentar cómo la reflexión ocupa un puesto irremplazable para escoger aquello que se vaya a eliminar y lo que vaya a permanecer.

Nos acercamos a textos ya constituidos. Nuestra dificultad se sitúa en encontrar unos procedimientos de solución tendentes a comprender y representar, más que a constituir novedosamente una síntesis creadora. El concepto de *interpretación* se ve aquí superado por el de *comprensión*. Debemos, primero, entender el texto, y luego construiremos una representación temática del mismo para la memoria inmediata, que tiende a desaparecer; mientras que la memoria a largo plazo, de uso más general y semántico, marca la primera macroestrategia: conservar en el recuerdo, salvar del olvido. Esta es la finalidad de nuestro proceso, y debemos conseguirla de la manera más económica posible.<sup>2</sup>

Un receptor documental manobra combinando información procedente de los propios hechos (información interna: contenido semántico y estructural); a lo que añade la situación contextual (información interna/externa: tipo de documento, actitud del autor y del receptor); y finalmente elabora unas presuposiciones cognitivo-esquemáticas (información externa). El carácter social marca las suposiciones contextuales: los hechos se comunican en un determinado entorno social, por lo que están dotados de funcionalidad intencional. No se limitan a ser estructuras lingüísticas. Surgen como actos expresivos pragmáticos.

<sup>2</sup> Véase a estos extremos H. J. Eikmeyer: «Procedural Analysis of Discourse», *Text*, 1983, 3, pp. 11-38.

Así pues, la lectura se produce de manera integradora: para entender una idea hay que comprender cada uno de los elementos que la componen. Desde la psicología cognitiva, los conceptos se consideran según sea su posición en los esquemas de conocimiento, por lo que son de una enorme complejidad:<sup>3</sup>

1. Esquemas de conocimiento:
  - 1.1. Experiencias anteriores (hechos, actos, causas, fines = enlaces pragmáticos).
  - 1.2. Conocimiento de los discursos (enlaces sintácticos y superestructurales).
  - 1.3. Conocimiento especializado (enlaces semánticos).
  - 1.4. Además, la información nueva o remática que el texto aporta.
2. Contexto sociocultural.

Vemos cómo el lector interpreta la situación que el texto le ofrece desde su experiencia sintáctica, semántica y pragmática. Efectúa saltos continuos desde su memoria al modelo de situación y al texto (*top-down processing*).<sup>4</sup> Este complicado proceso que permite tanto escribir como interpretar un texto se define mediante las estrategias, que son actuaciones que explican cómo utilizan el texto tanto el autor como el receptor, por más que las estrategias de producción y comprensión de los textos solo coincidan en los niveles de abstracción textual más elevados, diferenciándose notablemente en la operatividad concreta de acuerdo con sus objetivos divergentes. Se aprecian así los procesos retóricos como estrategias para alcanzar la estructura de las partes mayores del texto, o su totalidad.

El esquema retórico de la producción nos explica las operaciones y los componentes estructurales teóricos<sup>5</sup> en los que podemos diferenciar unos procesos lingüístico-textuales y unos procesos particulares que marcan la realidad de la comunicación, los textos concretos. Así, por la *dispositio* o *superestructura* el texto se fragmenta en unidades. La división que desea para su mensaje la consigue bien con variaciones de los campos léxicos, bien a través de esquemas novedosos. El documento nace para cubrir un determinado fin. El autor se vale de unos medios

<sup>3</sup> D. Rumelhart: «Schemata: the building blocks of cognition», en R. Spiro (ed.): *Theoretical issues in reading comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence and education*, Hillsdale: Erlbaum, 1985, pp. 75-92.

<sup>4</sup> También las ciencias del conocimiento abordan el procesamiento holístico que se efectúa durante la lectura; véase M. Minsky: «A frame work for representing knowledge», en P. Winston (ed.): *The Psychology of Computer Vision*, Nueva York: McGraw Hill, 1975, pp. 217-228.

<sup>5</sup> A. García Berrio: «Lingüística, literalidad/poeticidad. (Gramática, pragmática, texto)», en 1616. *Anuario de la Sociedad Española de Literatura General y Comparada*, 1979, p. 157.

para conseguirlo. Las estrategias contienen un dispositivo de adecuación en cuyo uso coinciden todos los hablantes y que permite la comunicabilidad del texto. Las oraciones de este se unen mediante el conocimiento *léxico*, el *lingüístico* (sintáctico-semántico) y el *contextual*. En los trabajos científicos cabe hablar de un cuarto tipo de empalme, el *metodológico* o *de género*, que se obtiene a través de la organización superestructural y que define las distintas unidades significativas que los conforman. Mientras leemos un texto se suceden saltos constantes desde los indicios de la memoria al modelo de situación y al texto concreto que nos ocupa. Son los procesos inferenciales mediante los cuales comparamos lo que ya sabemos con lo nuevo que nos trae el texto (*top-down processing*). Vamos proyectando en el texto nuestros esquemas cognitivos (inferencias elaborativas), hasta alcanzar lo sustancial del mensaje (inferencias reductivas). A la par se va comprendiendo cómo está construido el texto (inferencias lógico-sintácticas) y se deriva información de aquellos conceptos que las palabras expresan (inferencias léxicas). Asimismo relacionamos la macroestructura con los términos, oraciones y frases del texto, con su contexto y con el conocimiento personal (inferencias lógico-semánticas y desde el contexto).

Debemos tener siempre presente que lo más normal es que la idea esencial se exprese al principio del texto (en el título y en la introducción) y que cada macroestructura parcial se explicita al principio del subtexto que vincula (a través de los subtítulos y de los párrafos iniciales de las partes, capítulos y secciones en que el texto se divide). El centro del texto contiene la argumentación; es donde se muestran las estadísticas, citas, mediciones, aplicaciones y ejemplos (conviene recordar que los ejemplos no se reproducirán nunca en el resumen ni en los índices, pero pueden ayudar a completar, precisar o matizar la idea de una sección del texto). En la parte final, en cambio, se presenta el balance, los resultados, las conclusiones. Es fundamental para nuestros fines que se alcance a representar el texto desde sus líneas principales. El resultado de la lectura será, por tanto, no solo la extracción de las ideas principales, sino su localización ordenada. Para demostrar esta disposición será conveniente ayudarnos de un esquema como paso previo a la redacción del resumen e incluso a la asignación de términos para establecer los índices.

Respecto al examen del documento debe señalarse que la precisión con que se puede examinar un documento depende en gran manera de su forma física; por lo que se pueden distinguir dos casos diferentes: documentos impresos y documentos no impresos.

1. Los documentos impresos constituyen el material habitual de las bibliotecas y centros de documentación cuyo fondo consiste principalmente en libros, revistas, informes, actas de congresos, etcétera. De forma ideal la comprensión completa de estos documentos requiere su lectura detallada. Sin embargo, *una lectura completa es a menudo impracticable y no siempre necesaria, pero el indizador debe asegurarse de que no ha descuidado ninguna información útil.*<sup>6</sup>
2. La consulta de los documentos no impresos exige una dedicación más detallada, debido a que sus contenidos pueden demandarse con tal concreción que, si no han sido descritos minuciosamente, la recuperación se torna inviable.

## 2.2. Estrategias que seguir en la lectura de los textos

Siguiendo una gradación en el proceso de reconocimiento del documento podemos fijar dos estadios:

1. En la primera fase del reconocimiento se desarrolla una *lectura de situación*, por la que nos orientamos en el entorno donde se ha producido el texto.

Según se va leyendo se plantean al texto unos interrogantes contextuales:

- ¿Conozco algo del autor, de su obra, de sus tendencias? Con estos planteamientos pretendemos situar el texto.
- ¿A qué tipo de documento pertenece el texto?, así como ¿qué orientación predomina en él: expone, analiza, juzga, describe? Esta tarea de identificación del texto tendrá utilidad después, cuando escribamos el resumen.
- A partir del título se inicia una búsqueda continua de la macroestructura: ¿cuál es el tema principal?, ¿vamos corroborando que se corresponde con el título?, ¿cuáles son los objetivos que el autor se ha fijado con la obra que nos entrega?

2. El reconocimiento intencionado de un analista documental podemos denominarlo *lectura activa*.

Veámos antes que, cuando el proceso de análisis lo efectúan personas, los conceptos principales de un documento se identifican mediante el análisis intelectual. La selección de estos conceptos surge de la comprensión global del texto, lo que no exige su lectura completa: es resultado de un examen «inteligente», pero no obli-

<sup>6</sup> Documentación. *Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización*. PNE 150-121. ISO 5963-1985.

gatoriamente exhaustivo del texto. Un texto se estructura en partes, que no nos pueden pasar inadvertidas en la representación documental. Los lugares más ricos para obtener información tanto de la macroestructura global como de la superestructura (y por tanto de las macroestructuras parciales) se han hecho depender de la terminología. Fueron determinados por Anderson<sup>7</sup> (coincide con la norma ISO 5963-1985), que marcó así las zonas del documento que merecen una mayor atención:

1. El título del documento, ya que lo esencial debe describirse en él. Su relevancia se origina en que aparece como representación explícita de la macroestructura global: los términos que conforman el título deben garantizarnos la importancia de los conceptos que denominan. Fue esta la razón de que se escogiesen los títulos como *corpus* terminológicos en los primeros programas de indización automática (para elaborar índices KWIC y KWOC).
2. Los subtítulos: sumarios, tabla o índice de materias. Al recoger los títulos parciales describen lo esencial de cada una de las divisiones del texto. Explicitan en este caso las macroestructuras parciales en disposición ordenada o, lo que es lo mismo, la superestructura del original.
3. El resumen de autor, cuando exista. Su presencia en los originales facilita enormemente el análisis y la representación de los documentos. El resumen analítico sería suficiente por sí mismo como subtexto de consulta tanto para reconocer los conceptos que representaríamos en la indización como para ser reutilizado a fin de confeccionar el resumen informativo. Es así por contener necesariamente la macroestructura global y la superestructura (es decir, las macroestructuras parciales del texto ordenadas), y además explicadas.
4. La introducción y 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 o cursivas. Para indizar y resumir no sirven ni las generalidades ni lo demasiado específico.

Para descubrir el contenido conceptual global del documento es preciso, por tanto, determinar el tema principal, así como los elementos más ricos en infor-

<sup>7</sup> M. D. Anderson: *Book indexing*, Cambridge: Cambridge University Press, 1971.

mación. A lo largo de todo el proceso se trata de lograr una imagen lo más exacta posible del documento original para, luego, en la representación poder alcanzar el deseable homomorfismo. El indizador debe retener todos los términos útiles que le permitan describir los conceptos y objetos que abarca el documento en la medida en que estos interesen a los usuarios de determinada unidad documental.

Después de examinar el documento, el indizador deberá identificar las nociones que son elementos esenciales de la descripción del contenido. Los organismos que patrocinan la realización del análisis deben establecer los factores que se consideran importantes en el campo temático cubierto por los índices y por los resúmenes.

Los textos se construyen en torno a unas cuestiones clave, las cuales permiten también que se sigan esos mismos caminos para reducirlos a lo fundamental. Las preguntas *¿quién?*, *¿cuándo?*, *¿dónde?*, y sobre todo *¿qué?* y *¿cómo?*, *¿por qué?* y *¿para qué?* determinan con sus respuestas la identificación de la materia que estamos tratando y sus principales circunstancias. Desde ellas, y 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. Obedecen a los criterios de redacción informativa conocidos como *paradigma de Laswell* (*loi de Quintilien* para los tratadistas franceses), que ya eran conocidos en la antigüedad por haberlos propuesto Cicerón en *De oratore*, y que, de alguna manera, recogen también la gramática de casos de Fillmore.<sup>8</sup>

— *¿Qué?*: determina las materias y asuntos a los que atiende el 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..., etc.). 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.

<sup>8</sup> H. J. Lasswell: «Estructura y función de la comunicación en la sociedad», en M. de Moragas (ed.): *Sociología de la comunicación II. Estructura, funciones y efectos*, 4.ª ed., Barcelona: Gustavo Gili, 1994. La gramática de casos de Ch. J. Fillmore: «The Case for Case Reopened», en *Syntax and Semantics 8: Grammatical Relations*, Nueva York: Academic Press, 1977, pp. 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*.

- *¿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 este 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 (período) en el que se ejecuta la acción. Es el contexto cronológico o noción temporal de un contenido, distinta de 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 hasta conceptos mínimos, como sucede en el análisis de las fotografías.

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

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

Es curioso ver cómo coinciden estas categorías para el análisis de la información con otras originadas en conceptualizaciones bien alejadas. Podemos así considerar cómo se corresponderían estas preguntas con las facetas universales de Ranganathan, en método propuesto para analizar por facetas el contenido de los documentos e identificar así los conceptos:<sup>9</sup>

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

Esta estructuración de los mensajes por facetas puede esquematizarse de acuerdo con el siguiente esquema de funcionalidad:<sup>10</sup>

<sup>9</sup> S. R. Ranganathan: *Prolegomena to Library Classification*, Bangalore: Sarada Ranganathan Endowment for Library Science, 1957.

<sup>10</sup> Esta es la propuesta hecha por Isabel Ferin Cunha a partir de la gramática de casos desarrollada por Fillmore y Pottier, en su «Estruturação de vocabulário», en Grupo TEMMA: *Análise documentária. A análise da síntese*, Brasília: IBICT, 1989, pp. 63-87.

<b>Instrumento</b>	qué medio
<b>Agente</b>	quién ejecuta o sufre la acción
<b>Modo</b>	cómo se realiza la acción
<b>Locación</b>	tiempo, espacio y situación de la acción
<b>Resultado</b>	beneficio o beneficiario de la acción
<b>Finalidad</b>	intención de la acción

Podemos fijar las siguientes cuestiones en cuanto ejemplos de los criterios que deberíamos retener cuando efectuamos el análisis de documentos que apliquen métodos experimentales:

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

Por ejemplo: aproximarse a los estudios de documentación desde la programación neurolingüística.<sup>11</sup>

### 2.3. Algunas recomendaciones para la reducción del texto

Lectura y análisis deben hacerse con atención, ya que es preciso darse cuenta de las palabras de *enlace* y de los *enlaces lógicos* entre las ideas del texto. Debemos, pues, fijarnos en cómo se empalman las ideas secundarias con la idea principal:

- enlaces lógicos  
*mediante*
- enlaces explícitos (a través de palabras).

<sup>11</sup> UNE 50-121-1991 *Documentación: Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización*, Madrid: AENOR, 1991.

Este acoplamiento de ideas se deberá respetar en el resumen y en la indización (para salvar la superestructura y la cohesión del original).

Resulta muy útil que, una vez terminada la lectura analítica, elaboremos un esquema con el plan general del texto. En él se ordenarán las ideas más representativas, uniéndolas mediante nexos lógicos.

Para identificar los conceptos fundamentales a los que quedará reducida luego la representación del texto se puede aconsejar esta serie de tácticas:

- Comenzar por preguntarnos: ¿cuál es el tema principal?, ¿se corresponde con el título?
- ¿Podríamos darle al documento un título alternativo?
- ¿Cuáles son los objetivos fijados por el autor?
- ¿Qué orientación predomina en el texto: expositiva, analítica, descriptiva, crítica, etcétera?
- Describamos en cinco líneas el perfil del asunto, marcando la división temática que muestre el texto. A veces esta vendrá manifestada explícitamente mediante subtítulos. Otras será el propio documentalista quien se verá obligado a identificar los diferentes subtextos en que se divide la superestructura. Para ello tiene como apoyo el cambio subsecuente en los campos semánticos, coincidente con la progresión de una parte del texto hacia otra.
- Localizar las palabras clave de un texto para extraer la superestructura y seguir el razonamiento del autor. Identificar los puntos que contienen la idea principal de cada párrafo del texto.
- Luego, esquematicemos el plan del texto (puntos que contienen las ideas principales de cada párrafo).
- Mientras consultábamos el texto ¿hemos subrayado algo? En especial, ¿necesitaba explicación alguna expresión subrayada?
- Buscar las palabras-locuciones de enlace; ¿hay enlaces lógicos?

Finalmente, debe considerarse que debemos:

- 1) indicar el resultado y omitir cómo se consigue;
- 2) para hechos conocidos, evitar ejemplos, definiciones y explicaciones o especificaciones adicionales;
- 3) eliminar las proposiciones con poco contenido;
- 4) omitir aquello que el autor juzgue menos importante;
- 5) prescindir de lo que resulte obvio, así como de puntos y argumentos marginales;

- 6) ser positivos, prescindir de lo que no se haya dicho o escrito;
- 7) tomar la clase, grupo o conjunto, no sus miembros.

Respecto a los enlaces explícitos, recordemos que un texto representa la progresión de un pensamiento (por eso es discursivo). Al leer hay que poner en claro las etapas de esta progresión. Esta misma utilidad tendrá cuando debamos redactar el resumen, ya que estos enlaces sintácticos nos ayudarán a dar cohesión al texto que creemos.

- Así, como punto de inicio (al principio), suelen aparecer estas locuciones:
  - *En primer lugar*, y similares.
  - *Por una parte..., por otra.*
  - *Al principio/para comenzar.* Verbos: *comenzar, iniciar*, o palabras de su familia.
- Dentro de la argumentación:
  - Es común añadir un segundo elemento de importancia igual, superior o inferior (de acuerdo con el contexto):  
*Así, después, pues, por añadidura, en segundo lugar, luego, además, también* (no a principio de frase); o con relación lógica: *(ahora bien, pues, así pues, por lo demás).*
  - O establecer comparaciones: *Así como, de la misma manera* (cuando aparezca mismo, parecido, semejante...)  
*Igualmente, simultáneamente, paralelamente.*  
*Otro tanto, lo mismo.*
- Para la conclusión:
  - *Así, de este modo, entonces..., así pues, por tanto, luego.*
  - *Por, en consecuencia.*
  - *En conclusión, como conclusión, para concluir.*
  - *En fin, a fin de cuentas, finalmente, para finalizar.*
  - *En suma, en resumen, abreviando.*

Estos son ejemplos de criterios susceptibles de aplicación en general para los campos científicos; en disciplinas particulares puede ser necesario formular otras cuestiones.

### Capítulo 3

## EL PROCESO DE INDIZACIÓN Y SU RESULTADO: LOS ÍNDICES

### 3.1. Concepto de indización y su procedimiento

Dentro del análisis documental, la indización es la operación que busca representar tanto a los documentos como a las preguntas de los usuarios mediante los conceptos más relevantes de su contenido. Por medio de la indización se identifican los conceptos (a veces solo el léxico) macroestructurales de cada documento, con el fin de establecer una correspondencia con la demanda informativa del usuario, frecuentemente muy concreta, para que pueda aprovechar al máximo la información proporcionada por un texto, un documento gráfico, un repertorio e incluso todo un depósito documental. La transcripción de estos conceptos suele hacerse utilizando términos o códigos incluidos en los lenguajes documentales.

De acuerdo con los principios de indización defendidos por UNISIST, la indización es la operación que consiste en «describir y caracterizar un documento con ayuda de representaciones de conceptos contenidos en el documento; es decir, transcribir en lenguajes documentales los conceptos tras haberlos extraído del documento mediante un análisis».<sup>1</sup> Debemos entender que la indización mediante lenguajes documentales no es necesario aplicarla para una pequeña colección de documentos, como puede ser una biblioteca de casa. La mayor cantidad de documentos y, por tanto, de contenidos, implica el uso de técnicas de indización más profundas y complejas.

Como herramienta de trabajo la indización debe permitir la recuperación y localización de informaciones útiles. Por esta razón se erige en actividad principal

<sup>1</sup> UNESCO: *Guía para la redacción de artículos científicos destinados a la publicación*, París: UNESCO, Programa General de Información y UNISIST, 1983, p. 5.