

Capítulo 1

CONCEPTO Y CARACTERÍSTICAS DEL LENGUAJE DOCUMENTAL

El objetivo fundamental de todo centro de documentación o biblioteca es facilitar la recuperación de documentos o de la información contenida en los documentos. Dicha recuperación puede llevarse a cabo por diferentes vías: a partir del nombre del autor de la obra, del título o del tema que trate.

La información que ofrecemos en este manual va destinada a la recuperación en base al contenido conceptual del documento. El procedimiento seguido para conseguirlo con eficacia se denomina indización, que consiste en el análisis e identificación de los conceptos del documento, la selección de aquellas nociones que representen con mayor fidelidad la información que contiene y su traducción a un lenguaje documental.

La principal característica de los lenguajes documentales es el control terminológico, que permite la representación de los conceptos derivados de la indización sin ambigüedades, garantizando el empleo de los mismos términos para expresar los mismos conceptos.

1. CONCEPTO DE LENGUAJE DOCUMENTAL

Es posible definir el lenguaje documental desde un punto de vista exclusivamente pragmático, como útil al servicio de la recuperación de información, pero hay que considerar que en su composición intervienen elementos teóricos, pragmáticos y normativos que deben estar presentes en su conceptualización. Hay que tener también en cuenta que, por tratarse de una disciplina en formación, está sometida a los cambios inherentes a su evolución.

Podemos considerar el lenguaje documental como **todo sistema artificial de signos normalizados, que facilitan la representación formalizada del contenido**

de los documentos para permitir la recuperación, manual o automática, de información solicitada por los usuarios.

Las primeras manifestaciones del lenguaje documental datan de finales del siglo XIX, cuando aparecen las clasificaciones bibliográficas. Estas clasificaciones, inspiradas en la lógica y en los sistemas filosóficos del conocimiento, se fundan en el principio de precoordinación y son de carácter enciclopédico.

El concepto moderno de lenguaje documental se consolidó en el siglo XX cuando Cutter introdujo el lenguaje de encabezamientos de materia, basado en los principios de especificidad y de entrada directa, que señala el comienzo del desarrollo de lenguajes documentales especializados.

Los lenguajes especializados nacieron como respuesta a la dispersión del conocimiento, que obligó a una creciente especialización de los contenidos. Se crearon múltiples tesauros sectoriales, que resolvieron el problema de organizar los fondos documentales a escala institucional, pero esa misma diversificación trajo consigo el problema de compatibilidad entre los lenguajes y, como consecuencia, entre los centros de documentación y bibliotecas que los utilizaban.

Algunos especialistas son de la opinión de que en el proceso histórico evolutivo se tiende de nuevo hacia el enciclopedismo temático, lo que podría significar un movimiento cíclico en la evolución de los lenguajes documentales.

2. CARACTERÍSTICAS DEL LENGUAJE DOCUMENTAL

Previa consideración de las posibles relaciones del lenguaje documental con determinadas disciplinas, examinamos sus principales características, que permiten advertir las bases pragmática y normativa en las que se asienta.

El lenguaje documental es considerado como objeto de estudio de la lingüística documental, cuyas bases teóricas proceden de ramas del conocimiento tales como la lingüística general, la semántica, especialmente la estructural, la sintaxis o gramática, la terminología u onomasiología, la lexicología y la lingüística textual, entre otras. Esta cualidad del lenguaje documental de objeto de la lingüística documental es, sin duda, un fundado argumento en favor del establecimiento de esta como disciplina autónoma, al igual que el hecho de que tenga su campo definido, sin embargo, desearía poner de manifiesto mi falta de convencimiento en cuanto a la categorización de la lingüística documental como disciplina independiente.

Por una parte, el hecho de que el lenguaje documental es anterior al nacimiento de la lingüística documental como disciplina; a este respecto, recordamos las palabras de Otlet cuando escribía que *no se debe confundir el objeto de una ciencia con la ciencia que trata de ese objeto*.¹ Por otra parte, no parece posible considerar la lingüística documental, que no es lingüística aplicada y que tiene como campo de estudio el lenguaje artificial, como una rama en formación de la lingüística general, que se ocupa del lenguaje natural.

El lenguaje documental, objeto o no de la lingüística documental, tiene conexiones con diversas ramas del conocimiento, entre otras: con la inteligencia artificial, la lógica, y, en la medida en que el proceso documental es un proceso de comunicación, y el lenguaje documental es mediador en ese proceso, se le pueden atribuir relaciones con la teoría de la comunicación y con la teoría de la información. Es evidente, por otra parte, la existencia de influencias de las especialidades mencionadas al comienzo del epígrafe (semántica, especialmente la estructural, sintaxis, lingüística textual, etc.) con el lenguaje documental. De entre ellas destacaría las que ejercen el estructuralismo lingüístico y la lingüística textual.

Existen conexiones entre el lenguaje documental y el estructuralismo lingüístico en la medida en que este se ocupa de las unidades mínimas de la lengua, interesándose más por las relaciones existentes entre ellas que por esas unidades en sí mismas. Siendo el lenguaje documental concebido como conjunto de unidades significativas organizadas bajo determinada estructura, es obvio que el método estructuralista puede ser utilizado para la creación de relaciones artificiales entre los términos componentes de ese lenguaje.

A finales de los años 60 surge la lingüística textual, también denominada ciencia del texto o análisis del discurso, que supone un cambio total en las concepciones del análisis del lenguaje en su aspecto semántico. Sus precursores Petöfi y Van Dijk² centran la atención en el análisis global del texto, a diferencia de las gramáticas precedentes, dedicadas al estudio de unidades menores: palabras o frases.

La principal característica de la ciencia del texto es que respeta el análisis del documento, no se limita a los aspectos lingüísticos, sino que tiene en cuenta la materia de referencia, de ahí su interdisciplinariedad. El análisis documental, que comparte estas características, tiene evidentes connotaciones con la lingüística textual: en ambas el objeto de análisis es tanto la macroestructura del documento,

¹ P. Otlet: *Traité de Documentation*. Bruxelles: Editions Mundaneum, 1934, p. 10, epígr. 115.5.

² T. Van Dijk: *La ciencia del texto: un enfoque interdisciplinario*. Barcelona: Paidós, 1983.

como sus microestructuras; las dos disciplinas se sirven de la lengua para la desestructuración textual, con la diferencia de que la ciencia del texto se sirve del lenguaje natural y el análisis documental utiliza el lenguaje artificial para representar y recuperar la información pertinente del texto, lenguaje que al propio tiempo tiene su origen en ese proceso de análisis.

2.1. Partes componentes del lenguaje documental

El lenguaje documental es un lenguaje no natural, aunque utiliza los signos de este. Esos signos adquieren valor semántico por medio de su normalización y de las reglas morfosintácticas que lo articulan. Son precisamente esas reglas las que le dan categoría de lenguaje.

Los signos del lenguaje natural son las palabras, que representan nuestro conocimiento de la realidad, el lenguaje documental se sirve de ellas y, en ocasiones, las reemplaza por símbolos cargados de significado preciso de forma que lleva a cabo la representación del documento en virtud de una correspondencia analógica. Esta representación convencional tiene lugar cuando simboliza los elementos constitutivos del lenguaje de descripción (metalenguaje) por medio de instrumentos visuales de representación —la notación simbólica—; este es el caso de los sistemas de clasificación, que por el procedimiento de notación representan ideas de manera unívoca a través de números ordinales.

2.2. Principios del lenguaje documental

En razón de su propia naturaleza, el lenguaje documental no puede ser redundante. Si consideramos además sus fines, comprendemos que debe estar dotado de una sencilla organización.

El lenguaje documental, surgido de un proceso selectivo de la documentación, no deja lugar para la reiteración de términos, su eficacia aumenta a medida que la redundancia es eliminada. Por todo ello, este lenguaje debe regirse por el principio de entropía que, como concepto opuesto a la redundancia, se corresponde con la mayor cantidad de información aportada por un solo signo o mensaje. Precisamente el carácter entrópico del lenguaje documental es lo que le diferencia del natural, que, por el contrario, es asimétrico, porque la lengua es fundamental-

mente oral y requiere mucha redundancia. La reiteración de significantes se justifica aquí por el hecho de que facilita la recepción de los mensajes, a pesar de la interferencia de los ruidos. El lenguaje controlado tiene como misión suministrar la redundancia del lenguaje natural, lo que hay que tener presente en el momento de su elaboración. El lenguaje documental ha de ser unívoco, por ello se llama controlado, no puede permitirse la ambigüedad del lenguaje natural.

En comparación con el crecimiento de las ideas, el desarrollo del lenguaje natural es lento, como consecuencia, es necesario utilizar la misma palabra para expresar dos o más ideas. Ello da lugar a la homonimia: el mismo término se utiliza en diferentes sentidos en diferentes disciplinas, lo que produce distorsiones en la comunicación. La sinonimia es otro accidente lingüístico habitual en el lenguaje documental, que debe ser reconocido y, o bien eliminado, o, preferentemente, controlado.

Derivados de esos accidentes surge la ambigüedad, que crea dificultades estrechamente relacionadas con la aplicación de un lenguaje artificial a la recuperación de documentos. Se trata del ruido y silencio documentales, que se oponen a la relevancia de la indización.

Cuando se prepara una frase documental con objeto de recuperar un conjunto de documentos que contienen la información expresada en la frase, aquellos que no se recuperan, aunque existan en el archivo, son los productores del silencio documental, mientras que los que se recuperan sin haber sido solicitados —porque responden solo parcialmente al perfil de búsqueda— constituyen el ruido documental. Solo el logro de la deseable entropía puede evitar estos inconvenientes y facilitar una recuperación eficaz de la información.

2.3. Objetivos y funciones del lenguaje documental

El lenguaje documental interviene en dos fases del proceso documental, en el momento de la descripción y en el de la recuperación de la información. El objetivo de dichas operaciones es el de facilitar la recuperación de la documentación reduciendo el esfuerzo y gasto de tiempo del usuario. El lenguaje documental tiene capacidad para representar los mensajes contenidos en los documentos, lo que le permite cumplir dos objetivos fundamentales en el proceso, el de normalización y el de inducción, estando encaminadas a este último todas las demás funciones que desempeña a lo largo del proceso documental.

El lenguaje documental reduce considerablemente el volumen de términos del lenguaje natural, no tomando en consideración más que los sustantivos o los sintagmas nominales, seleccionando además, de entre ellos, un solo término entre dos o más nociones consideradas sinónimas. En virtud de estas características, se le reconoce un carácter mediador en los procesos de resumen e indización al controlar la posible ambigüedad entre términos.

En la fase de descripción, tras la lectura e identificación de los conceptos contenidos en los documentos, se representan mediante un lenguaje documental, que proporciona un vocabulario unívoco que favorece la coherencia del análisis y permite traducir los conceptos en términos normalizados.

Su objetivo primordial se cumple, sin embargo, en la segunda fase del proceso, suministrando el vocabulario de los conceptos de la demanda. El lenguaje documental proporciona una base para efectuar búsquedas al nivel de generalidad o especificidad deseados. En este sentido, se le puede considerar un lenguaje intermediario o metalenguaje en la medida en que sirve de puente entre las informaciones contenidas en los documentos y las informaciones solicitadas por los usuarios. Van Slype³ considera que los lenguajes de indización pueden intervenir en seis momentos diferentes del proceso de búsqueda:

- Selección de los sistemas documentales que se van a interrogar.
- Enunciado de los conceptos de la pregunta, en lenguaje natural.
- Traducción a un lenguaje de indización.
- Formulación de la ecuación.
- Extensión asistida por el ordenador.
- Apreciación final de la pertinencia.

Además de las mencionadas funciones, el lenguaje documental es de gran utilidad para la ordenación o archivo de documentos. Resuelve igualmente problemas planteados por el multilingüismo, como en el caso de los tesauros multilingües, que sirven de puente a diferentes lenguas, o en el caso de los lenguajes de clasificación, fundados en la simbolización numérica, que los hace independientes de toda lengua y constituyen una escritura comprensible en todos los idiomas, de ahí su fácil empleo internacional.

³ G. van Slype: *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Madrid: Fundación Germán Sánchez Ruipérez, 1991, p. 161.

3. TIPOLOGÍA DEL LENGUAJE DOCUMENTAL

La variada tipología del lenguaje documental le convierte en elemento de apoyo a disciplinas como la biblioteconomía, documentación, bibliografía y archivística, para cuyas necesidades de descripción ofrece posibilidades concretas. En relación con el análisis formal, el lenguaje documental completa el proceso técnico de catalogación dotando al soporte de la descripción de puntos de acceso temáticos.

Existen diversos criterios de tipificación de los lenguajes documentales, los más generalizados son: el de **control**, el de **coordinación** de los términos y el **estructural**.

Dependiendo del **control** ejercido sobre el vocabulario, los lenguajes pueden organizarse en dos categorías: *libres* y *controlados* (Figura 1).

Control	Libres	Listas descriptores libres
		Controlados

Existe abundante literatura acerca de las ventajas y desventajas que conlleva el uso del lenguaje libre y del controlado. Del análisis comparativo de uno y otro podemos concluir que el lenguaje controlado neutraliza las deficiencias del lenguaje libre y viceversa, por ello muchas bases de datos combinan la utilización de ambos en las distintas fases del tratamiento documental.

Los lenguajes libres, fundados en el principio de postcoordinación, se componen de un vocabulario no predefinido que se va generando a partir de la realización de procesos de indización. A excepción de las listas de descriptores libres y las de palabras clave, los demás lenguajes entran dentro de la categoría de controlados, que presentan un vocabulario previamente elaborado y que admite un limitado número de modificaciones en el momento de su utilización.

La sistematización de los lenguajes documentales según el criterio de **coordinación** (Figura 2) se realiza en función del momento en que se combinan los elementos que los componen. Si los términos se combinan cuando se elabora el lenguaje o en el momento de la descripción, el lenguaje será **precoordinado**, y si lo hace en el momento de la recuperación, se tratará de un lenguaje

postcoordinado. Van Slype, con un planteamiento pragmático, denomina lenguajes de clasificación a los precoordinados y lenguajes de indización a los postcoordinados; ambos tipos representan el contenido del documento de forma sintética y analítica, respectivamente.

Coordinación		
	Precoordinados	Clasificaciones
		Listas de encabezamientos de materia
	Postcoordinados	Listas de descriptores libres
		Listas de palabras clave
		Tesauros

En los sistemas precoordinados los términos están unidos de acuerdo con unas reglas sintácticas establecidas por el propio lenguaje. Los postcoordinados carecen de sintaxis; salvo excepciones, su vocabulario consiste en términos simples o unitérminos que se pueden conectar en el momento de la recuperación por medio de una sintaxis booleana. Atendiendo a los procedimientos seguidos para relacionar los conceptos al recuperar la información, podemos concluir que en los lenguajes precoordinados la relación entre los términos es gramatical, mientras que en un lenguaje postcoordinado la relación es lógica.

Los lenguajes precoordinados contribuyen mejor que los postcoordinados a conseguir precisión en la búsqueda, es decir, tienen mayor capacidad para rechazar los documentos irrelevantes en el momento de la recuperación porque las relaciones que se crean mediante el álgebra de Boole son genéricas y no impiden las falsas combinaciones. Como contrapartida, los postcoordinados tienen la ventaja de dar respuesta a necesidades de indización reales.

Combinando las características de control y coordinación, Courier⁴ reconoce la existencia de cuatro tipos principales de lenguajes documentales (Figura 3).

	Precoordinados	Postcoordinados
Libres	Sistema Precis	Sistema Unitérmino
Controlados	Clasificaciones	Tesauros
	Listas de encabezamientos de materia	

En cuanto a la **estructura**, se reconoce la existencia de dos tipos: **jerárquica** y **combinatoria**. A la estructura jerárquica responden determinadas clasificaciones, a la combinatoria, los léxicos documentales y los tesauros, si bien es cierto que en la organización multidimensional de estos últimos participa también la estructura jerárquica —en concreto en la presentación sistemática—. Presentados de forma esquemática, los lenguajes documentales, en función de su estructura, pueden ser:

Estructura	Jerárquica	Sistemas de clasificación jerárquicos
	Combinatoria	Léxicos documentales
		Tesauros

3.1. Definición de los distintos lenguajes

Damos a continuación un repaso general a las características y funcionamiento de los distintos lenguajes documentales y en los siguientes capítulos se tratan aquellos más significativos dentro de cada uno de los tipos mencionados anteriormente; se presta mayor atención, no obstante, a los que se utilizan más comúnmente en nuestra área geográfica: las listas de encabezamientos de materias, la Clasificación Decimal Universal y los tesauros. El uso combinado de los dos primeros procedimientos beneficia las tareas de clasificación y recuperación, por ello, se utilizan juntos en bibliotecas. Los tesauros combinan las características de los sistemas anteriores, al concebir en su estructura índices alfabéticos y sistemáticos.

⁴ Y. Courier: *Analyse et langage documentaires. Documentaliste*, v. 13, n° 5-6 (1976), p. 183.

LISTAS DE PALABRAS CLAVE

Listas no estructuradas de conceptos ordenados alfabéticamente, seleccionados por métodos automáticos, para describir el contenido de los documentos analizados. El método de determinación de las listas es singular en cada documento de modo que no permiten la descripción de otros materiales.

Los partidarios de la utilización del lenguaje libre —Chastinet y Salton, entre otros— apuntan la ventaja de su valor informativo intrínseco: se trata a menudo de términos muy específicos y actuales que permiten seguir de cerca la evolución terminológica de un dominio en particular. A esta ventaja añaden otras como el ahorro que supone no tener que elaborar y mantener un lenguaje documental, etc. Sus detractores señalan como principal desventaja la ausencia de control eficaz del vocabulario a causa de la ambigüedad del lenguaje natural. El usuario está sometido a las variaciones gramaticales de un mismo término y a las expresiones sinónimas de un mismo concepto.

SISTEMAS DE CLASIFICACIÓN

Presentan una distribución sistemática de conceptos en diversas categorías o clases, de tal manera que cada materia tiene un lugar predefinido. Se trata de un lenguaje codificado de forma numérica, alfabética o alfanumérica que pretende ser la descripción sintética del contenido de los documentos. Su utilización es habitual en bibliotecas por su carácter enciclopédico.

Se le reconoce como inconveniente su falta de operatividad por la rigidez de su estructura.

LISTAS DE ENCABEZAMIENTOS DE MATERIA

Concebidas en función de búsquedas secuenciales en archivos manuales e informatizados, presentan una estructura asociativa o combinatoria.

Un encabezamiento de materia trata de representar el o los temas de un documento con la ayuda de una serie de términos expresados en lenguaje natural y relacionados entre sí. El encabezamiento procede, por lo tanto, de la precoordinación de términos en vista de búsquedas en los ficheros alfabéticos de materias. De hecho, el uso de las listas de encabezamientos se materializa en los catálogos diccionarios que se utilizan en bibliotecas, en donde se realiza una práctica ge-

neralizada: la combinación de una clasificación y un catálogo alfabético de materias. Las bibliotecas pueden concebir su catálogo de tal manera que concuerde con su sistema de clasificación.

TESAUROS

Lenguaje postcoordinado de estructura combinatoria, está constituido por listas estructuradas de conceptos —descriptores— que posibilitan la descripción analítica del contenido de los documentos.

El control terminológico en un tesoro es muy alto, y los descriptores se conciben de tal forma que los conceptos de los documentos pueden ser expresados independientemente los unos de los otros. Para evitar ambigüedad y alteraciones de sentido, el tesoro se forma en base a unitérminos y descriptores compuestos, ya que la univocidad de ciertos conceptos precisa que el sustantivo vaya acompañado por un adjetivo.

4. CUESTIONARIO DE AUTOEVALUACIÓN

1. ¿Qué es un lenguaje documental?
2. ¿En qué principios se fundamenta el lenguaje documental?
3. ¿Cómo adquieren valor semántico los signos que componen el lenguaje documental?
4. Señale los objetivos fundamentales del lenguaje documental.
5. ¿En qué fases del tratamiento documental resulta esencial el uso de un lenguaje documental?
6. Atendiendo al criterio de coordinación, ¿a qué tipo de lenguaje documental pertenecen las listas de encabezamientos de materia?
7. ¿En qué principio se funda el lenguaje libre?
8. En función de su estructura, ¿a qué tipo de lenguaje documental pertenecen los tesauros?
9. ¿En qué principio se fundan las clasificaciones bibliográficas?
10. ¿Por qué las listas de descriptores libres entran dentro de la categoría de lenguajes postcoordinados?