

CAPÍTULO II

BASE LÉXICA Y MACRORREGLAS

2.1 Precisiones sobre la base léxica

Un componente esencial de los lenguajes asociativos es la presencia de la lengua natural en la formación del vocabulario procedente de un discurso determinado. Este vocabulario o **base léxica** constituye el soporte sobre el que habrán de erigirse todas las construcciones. De alguna forma, el vocabulario determina la estructura del lenguaje y también el sistema de reglas. De hecho, a partir de un vocabulario resultará uno u otro lenguaje documental y esto se debe, de un lado, a la naturaleza y tratamiento de los conceptos cubiertos y, de otro, a la tipología de categorías léxicas incorporadas.

En cuanto a las características de los conceptos y sus representaciones, los descriptores, conviene consolidar la visión pragmática de los lenguajes documentales mediante la comparación con sistemas análogos o afines: léxico, vocabulario, nomenclatura y terminología. Para ello, nos basamos en Cintra et. al.¹ quienes, partiendo de Dubois², analizan esclarecedoramente las relaciones y divergencias de esos constructos respecto a los lenguajes documentales. Las autoras brasileñas se centran en la observación de los usos sinónimos de las cuatro denominaciones estableciendo, con precisión, sus denominaciones:

- **léxico**: designa el conjunto de unidades reales o virtuales que forman la lengua de una comunidad: elementos y reglas que permiten la construcción de nuevas unidades.
- **vocabulario**: conjunto de ocurrencias que integra un determinado corpus discursivo: el vocabulario empleado en un texto.

Para Cintra y sus colegas, los lenguajes documentales no son léxicos, en el sentido anterior de la acepción, al no estar suficientemente articulados ni constituirse en unidades generadoras de nuevos elementos (restricción parcialmente superada en los lenguajes que postulamos). Por otra parte, tampoco contienen vocabulario puesto que combinan palabras de distintos discursos y procedencias (textos, usuarios, terminologías, índices) para cumplir sus objetivos de regulación del análisis y la recuperación.

- **nomenclatura**: es una lista de nombres, como indica la propia palabra, que supone biunivocidad de la relación significante-significado. Sus unidades son etiquetas dadas a conceptos independientemente de un sistema notacional particular: oro, agua.
- **terminología**: conjunto de términos definidos rigurosamente en función de las nociones e intereses particulares de un área de conocimiento: terminología de la comunicación y cultura, por ejemplo.

Las autoras concluyen que los lenguajes documentales no se confunden con las cuatro definiciones previas aunque incorporan elementos de todas ellas³.

Realizada esta aclaración desde la lengua natural, es conveniente indicar que las unidades léxicas, sumadas a las categorías posibles en las que aquéllas se transportan, constituyen lo que llamamos base léxica de los LE: conjunto de descriptores e identificadores que proceden del léxico, de los vocabularios, de nomenclaturas y terminologías, entre otras fuentes y, además,

expresión sinónima, considerando las advertencias anteriores y si no se expresa particularmente, de léxico, terminología o vocabulario (de los mismos lenguajes).

2.2 Las categorías léxicas en los lenguajes epistemográficos

Anteriormente vimos como la investigación documentológica en Análisis y lenguajes documentales avanza poco desde el momento en que la norma frena la libertad de presencia y uso de categorías léxicas distintas al sustantivo en el vocabulario de un tesoro. En los LE, y en función de las características del discurso que pretende representar (en nuestro caso el área del Patrimonio histórico andaluz), debe admitirse un mayor número de categorías léxicas para poder dar adecuada cuenta de los contenidos de los documentos y de su localización precisa. Valga un ejemplo para clarificar nuestra postura: no es lo mismo, en un banco de imágenes patrimoniales, buscar fotografías del rey Alfonso XIII sentado en una visita a Sevilla o subiendo una escalera de los Reales Alcázares. La acción subir (o sentarse y muchas otras) no tiene correspondencia fidedigna en la categoría sustantiva en castellano que "objetiza" algunos movimientos y estados (asientos, subidas). Así, los descriptores que expresan "subir a un caballo" (la acción "subiendo" congelada por el obturador), en un tesoro, probablemente serían montería, cabalgada, cabalgadura o silla de montar y para "El rey sentado", asiento, sillón, sofá o directamente una ecuación en la que hacen presencia el

agente y el lugar: /rey/ /Sevilla//Reales Alcázares/. A nadie se le escapa la pobreza de estas representaciones y la cantidad de distorsión en la recuperación.

Estas lagunas léxicas, muy perjudiciales en un sistema de información visual o audiovisual, también afectan a la documentación textual cuando en la lengua natural no hay sustantivo que refleje el concepto contenido en adjetivos o verbos. En suma, la base léxica de un LE debe ser lo suficientemente amplia, en unidades y categorías, como para representar los temas esenciales de cualquier documento del área y cubrir las demandas sobre el mismo, articular nuevas unidades y enunciados y, más allá de la función instrumental, informar sobre fragmentos o porciones del discurso.

En consecuencia, si los problemas de formalización de los significantes comprometen los resultados del análisis y la recuperación documentales y podemos aprovechar el marco de las nuevas tecnologías y las transformaciones que introducen en el tratamiento de enunciados, consideramos que las siguientes categorías son ineludibles para transportar los significados en un sistema de información:

2.2.1 Sustantivos y sintagmas nominales

Los sustantivos han dominado, tradicionalmente, el ámbito de la representación desde que irrumpió en la escena documental la lengua natural, como sustituta de la codificación, a mediados de los cincuenta. La aparición, en los tesoros, de sintagmas nominales como descriptores compuestos no fue tanto una concesión a

los adjetivos, que debían siempre depender de una base sustantiva, como facilitar la jerarquización del vocabulario mediante simples precoordinationes:

- Instituciones
- Instituciones de enseñanza
- Instituciones de enseñanza superior

Bien es cierto que la precoordinación ha tenido como metas principales la eliminación de polisemias y la aceptación de expresiones consolidadas. La participación de adjetivos en los descriptores sintagmáticos alivia escasamente las retricciones porque, aun aportando al sistema la calificación esencial en un lenguaje, favorece el incremento desmedido de descriptores compuestos sin haber reglas precisas para señalar con nitidez la pertinencia o no de una composición.

No obstante, es obvio que un sistema de análisis documental necesita sustantivos para representar el universo conceptual, especialmente, aquello relativo a los objetos y contables pero también a incontables como las acciones. Es más, debemos abogar por el mantenimiento del sustantivo como categoría léxica de mayor utilización y estabilidad conceptual siempre que no fuerce o desvirtue el significado o implique la exclusión de otras categorías más ricas y ajustadas a determinadas representaciones.

2.2.2 Adjetivos

Los adjetivos arrastran un mundo de significación irrefutable y necesario en cualquier sistema de información. Su ausencia se debe a carencias tecnológicas y a tradiciones anquilosadas. Hasta el momento, las bases de datos se conforman con incluir adjetivos en los descriptores compuestos y recuperar por adjetivos mediante truncados: arquitectura efímera (*efímera), feria popular (*popular), mesa cuadrada (*cuadrada). Podemos advertir que cualquiera de los tres adjetivos utilizados (efímera, popular, cuadrada) recoge sentidos que modifican decisivamente los significados de los sustantivos arquitectura, feria y mesa y de otros muchos miles de sustantivos viables en el sistema. Pues

bien, la única forma admitida de usar los adjetivos sería como formantes de sintagmas lo que obligaría a realizar varios miles de construcciones, tantas como significados lógicos y pertinentes sean posibles en una área de conocimiento. No tendría justificación que /cuadrada/ sólo calificara a /mesa/ y no a silla, casa, plaza, o mantel y que /popular/ sólo adjetivara /feria/ y no barrio, mercado, televisión, cultura o fiesta. Si se opta por la ausencia se empobrece el vocabulario y si se admiten todas las construcciones el vocabulario crece exponencialmente lo que comprometería la gestión de la base de datos⁴.

En definitiva, parece que la única solución posible es permitir la participación del adjetivo aislado, como descriptores calificadores del lenguaje, en espera de construcción de un enunciado (mediante unión morfológica) por parte del usuario antes de acceder al sistema de información. Esto comporta una extensión ilimitada de combinaciones (marcadas únicamente por las fronteras discursivas) y una drástica reducción de la base léxica; es decir, disminuye el vocabulario y aumenta la capacidad de representación.

2.2.3 Verbos

Anatemizados por la normativa, los verbos, y las acciones que transportan son los grandes ausentes de la representación del análisis de contenido de las imágenes fijas y audiovisuales y de muchos textos escritos. A través de dos formas, gerundio y participio, podrían representarse todas las acciones y situaciones perceptibles en una imagen. Sin embargo, también esta laguna afecta al análisis de objetos y textos. Así, como veíamos, subiendo, saludando (transportable al sustantivo /saludo/), bajando, entrando (transportable a /entrada/ con la consiguiente polisemia) son categoría verbales -gerundios- esenciales para la descripción de una imagen en la que se percibe acción o movimiento. Por otro lado, sentado, roto, destruido, son participios que expresan estado o situación también perceptible en una imagen o denominan técnicas (barrido, ensamblado) descritas en textos o estado de los objetos (fragmentado, agrietado).

2.2.4 Otras categorías

Otras categorías y elementos léxicos, de rango menor, pueden ser necesarios como formantes de la base léxica de un lenguaje epistemográfico. En ese caso, el tipo de instrumento propugnado, asistido por la oportuna tecnología, asume sin ningún problema esas alternativas: los prefijos, que pueden ser muy habituales como formantes de término en un área determinada, deberían listarse aparte para ser llamados en caso necesario por su función hiponímica: mini, micro, super (ordenador), bio (Química, Tecnología, diversidad). Las desinencias de gerundio (ando, endo) y participio (ado, ido) podrían ayudar a la recuperación de acciones y situaciones expresadas visualmente a partir de peticiones de usuarios sobre descriptores y lexemas que sufrirían transformaciones morfológicas para seleccionar, con absoluta precisión, la imagen solicitada. Un listado de acciones/estados de carácter general debe ser parte del lenguaje que gestione este tipo de documentación. Determinados adverbios de lugar, modo o tiempo podrían también ser utilizados caso de no transportar, sus sentidos, las categorías básicas.

Una mención distinta merecen los identificadores que no son categorías sino elementos llenos del vocabulario. Los nombres propios de personas (reyes, santos, pintores, jueces), Instituciones o edificios (alhambra, alcázares) o proyectos (Unisist, Sócrates) no se incorporan habitualmente a los tesauros salvo como índices auxiliares. Según el tipo de discurso, los identificadores pueden adquirir una prevalencia, en análisis y demanda, que el documentólogo no puede ignorar. Los LE, de hecho, no realizan distinción en el tratamiento de denominaciones comunes o propias. Los identificadores dependen siempre de un descriptor genérico como enumeración o "ejemplos" del mismo pero cada uno de ellos desarrolla su propio escenario (pintores: Murillo, Picasso). En la aplicación que nos ocupa, debe advertirse que los identificadores patrimoniales pueden aumentar considerablemente lo que tendría una sola consecuencia: la necesidad de actualizar el lenguaje con mayor frecuencia.

Finalmente, encontramos, en un LE, construcciones léxicas artificiales forzadas a fin de organizar o dar cobertura a términos o expresiones afines: los **niveladores**. Su función primaria (nivelar o equilibrar) provoca, a veces, la creación de expresiones poco identificables y sólo entendibles y válidas en un contexto y acompañadas de notas que aclaren su valor convencional. A pesar de ello, la lógica de la poscoordinación genera una gran cantidad de niveladores muy útiles para montar la estrategia de análisis o búsqueda pero no utilizables para entrar en la base de datos.

2.3 Reglas morfosemánticas

Procede, a continuación, detenerse en las reglas de que dispone la base léxica, bien en su construcción inicial bien en su actualización o uso, estableciendo el punto de referencia en los recursos manejados por los tesauros. Número, género y composición aparecen como potentes instrumentos de la significación.

2.3.1 El número como significado

El rol del singular/plural en un lenguaje asociativo no es banal (plural=muchos=, singular=uno). En la recuperación de información, las referencias captadas por el descriptor contienen uno, varios o muchos elementos siempre con independencia del singular o plural de su morfología. Por ejemplo, /camión/ puede recuperar textos o imágenes que se refieran a o muestren uno o muchos camiones. A la inversa, puede ocurrir que /bomberos/ muestre un texto o foto referido a un individuo concreto. Es decir, el número de un descriptor no tiene correspondencia con el número de conceptos que representa el documento. Nuevamente encontramos la regla teórica frente a la realidad discursiva.

Esta salvedad es útil y aprovechable, puesto que si el número, en principio, no comporta significado previo, si no hay vinculación lógica entre el número en la representación y los elementos textualizados, este recurso lingüístico puede estar sometido a una macrorregla que



afecte a la generalidad o a partes del lenguaje. Habitualmente existen convenciones del tipo: contable=plural (camiones, bomberos), incontable=singular (erosión, técnica), entre otras.

Más allá de la regla adoptada, en los lenguajes epistemográficos, el número es un recurso esencial del control de polisemias a tenor de la ingente cantidad de descriptores necesarios que comparten significados distintos. Así, el estado plural de los lugares (cesterías, panaderías) genera una significación diferente de las mismas lexías en singular: las actividades con idéntica denominación, por ejemplo, al lugar en el que se produce o vende el producto (cestería y panadería serían acciones). Otro ejemplo diferencia, dentro de la misma macrocategoría -Agentes-, lo colectivo de lo individual: Policía nacional (singular=institución o colectivo), policías nacionales (plural=individuos). Obsérvese la aparente paradoja de usar el singular para un colectivo y el plural para individuos. En realidad, hay coherencia con la regla general pues se pluralizan personas contables mientras los entes son considerados únicos e incontables.

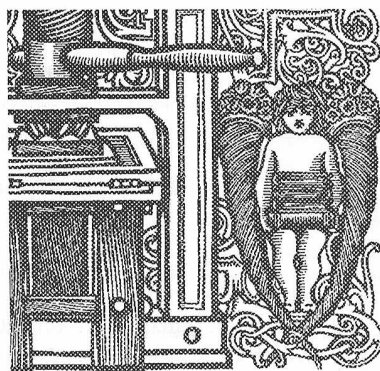
En suma, la presencia del número en la base léxica, lejos de ser casual o molesta, es consciente y beneficiosa. Permite extender el campo conceptual con un simple signo plural o su contrario en lugar de recurrir a sustituciones o sintagmatizaciones. Por ello, el número bien controlado y conocido por el usuario se convierte en una poderosa herramienta de clarificación y economía lingüísticas.

2.3.2 El género

Un lenguaje epistemográfico debe tener la suficiente flexibilidad como para permitir la representación con el descriptor más ajustado al concepto que cubre u ofrecer los datos factuales solicitados según el género. Por otro lado, no es de extrañar que desde la óptica femenina, la ausencia de su género en el vocabulario habitual de los thesauri provoque críticas fundamentadas.

A diferencia del número, que evita la poliseμία sin efectos secundarios, siendo el más habitual la inflación terminológica, la presencia de los dos géneros supone dos descriptores, uno para el valor femenino y, otro, para el masculino. En los ámbitos en los que sea preciso, no hay impedimento tecnológico alguno para mantener la dualidad.

Si no es necesario el mantenimiento de ambos géneros, como quiera que la cuestión tan sólo afecta al campo de Agentes (personas, roles, animales), muchos casos de la oposición a/o pueden ser absorbidos por el plural neutro -es: león/leona=leones, juez/jueza= jueces por lo que la cantidad de términos transcritos en ambos géneros es reducida. Afectaría a aquellos casos en los que las raíces son distintas (toro/vaca) o en los que el plural no es acompañado por la transformación vocal en "e" neutra (maestros/maestras). Una regla, para este grupo menor, puede zanjear el problema a fin de que los documentos y los usuarios coincidan en la etiqueta del género que refieren mientras el conjunto mayoritario se rige por la regla general.



2.3.3 Composición morfológica

Un lenguaje asociativo debe descansar en los enunciados generados por el análisis y por los usuarios, modelizados como ecuación de búsqueda, es decir, en los contextos. Este principio de la poscoordinación permite contar con un vocabulario simplificado, significativo y muy flexible. La tendencia de cualquier lenguaje documental, de las características mínimas de un tesoro, es el uso de unitérminos desambiguados o unisémicos. No obstante, los descriptores sintagmáticos son necesarios para evitar la ambigüedad de alguno de sus elementos antes de la composición: unión + monetaria, interés + bancario. Este recurso constructivo, la unión morfológica, está recogido en todos los manuales y normas sobre lenguajes combinatorios como un instrumento antipolisémico y, en todo caso, para ser usado en la precoordinación por los constructores del tesoro.

En los lenguajes epistemográficos, además de asumirse esa función, la composición morfológica es un recurso de la poscoordinación, materializándose como macrorregla del plano sintagmático o de las realizaciones enunciativas. Así, el usuario mediador o final dispone de la posibilidad de montar frases sintagmáticas con elementos provenientes de diferentes lugares del lenguaje, se convierte en verdadero constructor y actualizador del vocabulario (como el habla, para los estructuralistas, se construye a partir de una norma y un sistema). Esto ocurre muy particularmente, y con frecuencia notable,

a los adjetivos en busca de sustantivos para indizar o recuperar: rojo + estandartes = /estandartes rojos/, paredes + inclinado = /paredes inclinadas/; rojo e inclinado son adjetivos pertenecientes al campo /Atributos/ del TPHA, categoría especialmente afectada por las uniones morfológicas.

Debemos llamar la atención sobre el entrenamiento que han de seguir los usuarios para manipular o navegar en un lenguaje epistemográfico que recoge convenciones y recursos no habituales e innovadores. Partimos de la convicción de que, en poco tiempo, el usuario aprende y se adapta a los métodos de poscoordinación de la misma manera que lo hizo, en los años setenta, con los insuficientes y forzados, para el discurso cultural, operadores booleanos. La actual tecnología permite una mayor facilidad mediante el aprendizaje intuitivo de las reglas apoyado en sistemas tutoriales.

Es necesario, finalmente, establecer las debidas distancias entre composición morfológica en la poscoordinación y combinación sintáctica (habitualmente lógico-matemática): son recursos adscritos a distintos planos de actuación. Los primeros provienen de las potencialidades constructivas del propio vocabulario para crear nuevos conceptos, los segundos pertenecen a las reglas gramaticales del sistema informático y su función es construir oraciones.

2.3.4 Definiciones y notas

Como instrumento, un LE necesita las con-

vencionales notas de alcance -NA- de los tesauros, con sus funciones correspondientes:

- eliminación de polisemias, mediante la propia NA o un calificador: /ilustración/ NA: dibujos y gráficos.
- restricción o ampliación del significado: /trenes/ NA: sólo mercancías.
- aclaración de un término: /ciencia de la información/ NA: en América Latina equivale a la Ciencia de la Documentación en España.

Como producto, un LE proporciona la acepción de un descriptor desde el contexto terminológico y discursivo que da lógica a la base de conocimiento. Podríamos decir que el usuario dispone de una función terminológica de la base léxica: todos los descriptores van acompañados de una acepción determinada por el sistema nocional del que provienen. Ej.: Análisis de Contenido: procedimiento de detección y extracción de los temas o asuntos de un documento para su representación mediante palabras-clave, resúmenes u otros productos.

En el ejemplo citado, el valor asignado al descriptor /análisis de contenido/ es el uso que se le atribuye en el sistema nocional de la Documentología. Obviamente, /análisis de contenido/ tiene acepciones distintas en otros sistemas además de dar nombre a una Disciplina científica. Es importante señalar que el marco de la acepción es el contexto de la base de conocimiento. Así, la definición de /análisis de contenido/ como técnica documental sería la expresada en el ejemplo anterior, pero si /análisis de contenido/ existe también como disciplina, en el mismo sistema, habría que intervenir con dos medidas:

- impidiendo la polisemia con los recursos apropiados de tal forma que cada entrada refiera un solo significado y viceversa, por ejemplo, /análisis de contenido/ y /análisis de contenidos/
- definiendo, a continuación, los dos nuevos descriptores en su respectivos sentidos.

De lo expuesto se desprende una "función diccionario" o definitoria del LE aunque, a diferencia de las NA que aparecen espontáneamente, las definiciones se explicitan en pantalla a instancias del usuario.

2.3.5 Tratamiento de las equivalencias

La tecnología permite que el usuario acceda al vocabulario por cualquiera de las entradas de un concepto sin necesidad de decidir formas preferenciales sobre formas no autorizadas. Por tanto, los operadores de equivalencia, propios de glosarios, de listas de encabezamientos de materias y de tesauros, que remiten a las entradas permitidas (USE) o informan de las entradas no permitidas (UP, usado por) desaparecen como recursos de la base léxica de un LE. Deben tenerse en cuenta los siguientes factores en la escena terminológica de una epistemografía:

1. La omisión de los USE/UP no significa ausencia de control entre los sinónimos.
2. Debe incorporarse el mayor número posible de sinónimos lingüísticos (en sus acepciones especializadas) y reducir, al máximo, la permisividad de las cuasisinonimias en cualquiera de sus variantes: clasemática, lexemática y antonímica. Los cuasisinónimos empobrecen la base léxica mediante generalizaciones y, por tanto, perjudican la precisión de enunciados, escenarios, análisis, representación y búsqueda.
3. Todos los sinónimos, o entradas de un concepto, son válidos para acceder a la base de datos desde un LE. La centralización de cada sinónimo en un índice automático permite, al usuario, utilizar la entrada más familiar y, al sistema, comprender la demanda desde cualquier ángulo. Asimismo, el analista podrá representar el tema o concepto de un documento mediante la entrada que juzgue más adecuada.

Ejemplos

<u>Sinónimos de identificador</u>	<u>Concepto representado</u>
<ul style="list-style-type: none"> • Rey Juan Carlos • Juan Carlos I • Juan Carlos de Borbón 	(Juan Carlos I, rey de España)

<u>Sinónimos de descriptor</u>	<u>Concepto representado</u>
<ul style="list-style-type: none"> • medios de comunicación • medios de comunicación social • medios de comunicación de masas • mass media 	(medios de comunicación social)

Los operadores de equivalencia de los tesauros se sustituyen por una regla del sistema tutorial que advierte a los usuarios sobre el empleo de los sinónimos, truncados y otros procedimientos. La presencia de todos los sinónimos posibles (en función de un sistema notacional dado) en torno a un concepto desdobra la “función terminológica” de los LE en diccionarios de acepciones y diccionarios de sinónimos en el contexto del sistema.

NOTAS

1. Cintra, A. M.; Tálamo, F.G.M.; Lara, M.L.G. e Kobashi, N.Y.: Para entender as linguagens documentárias. Sao Paulo: Polis; Associação paulista de bibliotecários, 1994. 72 p.

2. Las autoras toman los rasgos distintivos del trabajo de Dubois, J. et al: Dicionário de Lingüística. Sao Paulo: Cultrix, 1973.

3. Cintra et al, op. cit., p. 25-27

4. Un diez por ciento de los términos del Tesauro de Patrimonio Histórico Andalúz son adjetivos adscritos al campo de Atributos (propiedades y características). La incorporación de adjetivos disminuye el vocabulario total en un 30% y potencia notablemente la articulación léxica naturalizando el lenguaje. Como contrapartida, el usuario debe hacer un mayor esfuerzo y estar familiarizado con el lenguaje y sus recursos.

5. Vid epígrafe 4.2.7.3 en el que se relacionan los campos del Tesauro según el número adjudicado.