

TESAUROS Y LISTAS DE EPÍGRAFES: ¿HACIA UNA INTEGRACIÓN?

[THESAURI AND LIST OF SUBJECT HEADINGS: TOWARDS AN INTEGRATION?]

ANA MARÍA MARTÍNEZ

Resumen: Se analizan las semejanzas y diferencias entre los tesauros y las listas de epígrafes, tomando en cuenta el plan de revisión 2003 de la norma estadounidense Z39.19. Esta norma pretende establecer nuevas directrices para todo tipo de vocabulario controlado, en relación con su utilización con nuevas tecnologías. Se concluye que, en efecto, es muy importante revisar toda la normativa referida a los vocabularios controlados, particularmente las de alcance internacional, pero al mismo tiempo se hace necesario tener sumo cuidado en mantener un riguroso control de la terminología, una correcta definición de las relaciones jerárquicas y asociativas y promover la poscoordinación gestionada, en lo posible, mediante sistemas facetados y computarizados.

Palabras claves: Tesauros; Listas de epígrafes; Normalización; Sistemas de organización del conocimiento.

Abstract: The similarities and differences between thesauri and the lists of subject headings are analyzed, taking into account the work plan for the 2003 revision of the US standard Z39.19. This standard pretends to establish new guidelines for all types of controlled vocabulary in relation to their use with new technologies. We conclude that it is very important to review all the standards for controlled vocabulary, particularly those with international scope, but at the

Departamento de Bibliotecología, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. 48 e/6 y 7, 1900 La Plata, Argentina. Correo electrónico: ammarti@netverk.com.ar

Artículo recibido: 22-09-03. Aceptado: 03-10-03.

INFORMACIÓN, CULTURA Y SOCIEDAD. No. 9 (2003) p. 9-27

©Universidad de Buenos Aires. Facultad de Filosofía y Letras. Instituto de Investigaciones Bibliotecológicas (INIBI), ISSN: 1514-8327.

same time, it is necessary to be very careful in maintaining a rigorous control of terminology; a correct definition of hierarchical and associative relationships, and to promote the postcoordination, managed, if possible, in faceted and computerized systems.

Keywords: Thesauri; Lists of subject headings; Knowledge organization systems; Standardization.

Introducción

Una de las oportunidades más interesantes que tienen los bibliotecarios ante sí en el ciberespacio es la de contribuir a la organización del conocimiento, no sólo en las bibliotecas virtuales sino también en sitios empresariales, gubernamentales, no gubernamentales, educativos, etc. Con este fin, se han probado con distinto éxito diversos sistemas de organización del conocimiento (SOC)¹: por un lado están aquellos de uso tradicional en las bibliotecas como los sistemas de clasificación, las listas de epígrafes, las listas de autoridades de materia y los tesauros. Por otro lado están los SOC de aparición más reciente y vinculados con el medio digital como las ontologías, taxonomías y redes semánticas (Hodge, 2000). No se ha dicho aún la última palabra en cuanto a la mayor o menor efectividad de unos u otros, pero algunos autores consideran que el tesaurus es uno de los SOC que mayores ventajas ofrece, ya que brinda acceso tanto verbal como sistemático a la información (Gilchrist, 2000).

Las listas de epígrafes o de encabezamientos de materia también han sido objeto de estudio en este sentido. Debido a que es un tipo de SOC que se inventó para el catálogo en fichas, durante la década de 1980, al instalarse definitivamente los catálogos en línea y comprobarse las serias dificultades que ofrecían para la recuperación por materia, aparecieron duras críticas a la lista de epígrafes *Library of Congress subject headings* (LCSH), que puede considerarse como el paradigma de este tipo de SOC (Bloomfield, 1993; Chan, 1989; Cochrane, 1984; Henige, 1987; Holley y Killheffer, 1982; Klugman, 1989; Reynolds, 1989; Shubert, 1991; Studwell, 1993; Studwell y Folsom, 1993).

En consecuencia, Library of Congress decidió tomar algunas medidas que garantizaran la supervivencia de esta lista. Entre otras cosas, en 1986 cambió sus tradicionales referencias de *see/x*, *see also/xx* (que en español serían véase/x, véase además/xx) por las referencias tesaurias *Use/UF*, *BT/NT* y *RT/RT* (*Use/UP*, *TG/TE*, *TR/TR*). Aunque el cambio pretendía lograr una mejor distinción entre las relaciones jerárquicas y las asociativas, LCSH no ha sido convertida realmente en un tesaurus y el análisis de estas relaciones entre epígrafes nunca se ha hecho a fondo (Bates, 1988; Dykstra, 1988; Rolland-Thomas,

1993). Así por ejemplo, *Diseases* (enfermedades) aparece como término específico de *Medicine* (medicina) que, desde luego, no es una relación jerárquica.

Sin embargo, la norma estadounidense para construcción de tesauros ANSI/NISO Z39.19 se hizo eco de este cambio ya en su revisión de 1993 (National Information Standards Organization, 1993), dando a entender veladamente la posibilidad de armonizar las listas de epígrafes con los tesauros, cuando afirmaba en el punto 2.1 que “la norma asume que los tesauros serán aplicados principalmente en sistemas de almacenamiento y recuperación de información poscoordinados. Por lo tanto, las directrices para el desarrollo de listas de subencabezamientos diseñadas para la precoordinación con descriptores no han sido incluidas en esta norma, *pero esto no evita el uso de tesauros en sistemas de indización precoordinados*”.

La misma National Information Standards Organization (NISO) ha encarado en 2003 una nueva revisión de la norma Z39.19, con el propósito de adaptarla a la creación de SOC apropiados a las nuevas tecnologías. La idea es ampliar el alcance de la norma, que actualmente se centra en la biblioteca y los documentos textuales e impresos, a otros ambientes como la World Wide Web y a otras formas de información, medios y formatos (National Information Standards Organization, 2003). Para ello se ha previsto el tratamiento de los siguientes aspectos incluidos en el plan de revisión de la norma que deja abierta la puerta para la inclusión de las listas de epígrafes, por ejemplo:

- Considerar otros tipos de vocabularios controlados, desde una simple lista de palabras claves hasta un tesoro completo, y brindar directrices y ejemplos para todo tipo de vocabulario controlado.
- En consecuencia, cambiar el término *tesoro* por el de *vocabulario controlado* a lo largo de toda la norma.
- Considerar la incorporación de vocabularios controlados precoordinados.
- Reducir la importancia de los calificadores y sustituirlos por términos compuestos.
- Promover la vinculación de los términos a las definiciones de un diccionario.
- Hacer que las formas gramaticales sean menos rígidas.
- Reducir el énfasis en las frases nominales.
- Considerar situaciones donde no sean necesarias todas las relaciones semánticas.
- Combinar las relaciones jerárquicas y asociativas en *categorías de navegación*.

En este trabajo se intenta reflexionar sobre estos aspectos, analizando las semejanzas y diferencias entre tesauros y listas de epígrafes y señalando sus ventajas y desventajas. En el análisis de los descriptores y sus relaciones se ha tomado en cuenta la norma ISO 2788 (International Organization for Standardization, 1986), así como otras normas nacionales y reconocidos manuales para la construcción de tesauros (Aitchison et al., 1997; Asociación Española de Normalización, 1989 y 1990; Instituto Argentino de Normalización, 1983; Lancaster, 1995; National Information Standards Organization, 1993). Para el análisis de los epígrafes se han tomado en cuenta varios autores (Chan, 1990; Rovira, 1966; Westby, 1984) y las *Directrices para los registros de autoridades de materia y sus referencias* de IFLA (International Federation of Library Associations and Institutions, 1995). Los ejemplos de descriptores han sido tomados del *Tesaurus de la Unesco* (Aitchison, 1995), mientras que los ejemplos de epígrafes pertenecen a la *Lista de encabezamientos de materia para biblioteca* (Biblioteca Luis Angel Arango, 1998) y a *Sears: lista de encabezamientos de materia* (Westby, 1984).

Semejanzas y diferencias en el vocabulario

Tanto los tesauros como las listas de epígrafes son vocabularios controlados que aceptan los términos de indización mencionados en la Tabla 1.

Términos	Tesauros	Listas de epígrafes
Términos de indización: nombres comunes	Descriptores o términos preferentes	Epígrafes o encabezamientos de materia
Términos de indización: nombres propios	Identificadores	Nombres propios como materia (de personas, dinastías o familias, entidades, reuniones y lugares) y títulos como materia
Subdivisiones comunes	No se aceptan	Subdivisiones o subencabezamientos
Términos no autorizados para la indización	No descriptores o términos no preferentes	Formas variantes

Tabla 1. Vocabulario de entrada de tesauros y listas de epígrafes

Existen desde luego diferencias terminológicas que deberían revisarse para lograr una terminología común (Martínez et al., 2003), pero conceptualmente la gran diferencia entre ambos es que las listas de epígrafes poseen subdivisiones y los tesauros no. Si la futura norma incluye a las listas de epígrafes es necesario establecer las directrices correspondientes a la redacción de estas subdivisiones.

En la Tabla 2 se presentan las formas gramaticales permitidas para los descriptores, comparadas con aquellas para los epígrafes.

Forma de los términos		Tesauros	Listas de epígrafes
Término simple - Un sustantivo		<i>Bibliotecas</i> <i>Organizaciones</i>	<i>Laboratorios</i> <i>Vuelos</i>
Término compuesto	Frase adjetivada	<i>Organizaciones culturales</i> <i>Organizaciones culturales internacionales</i>	<i>Vuelos espaciales</i> <i>Vuelos espaciales tripulados</i>
	Frase preposicional	<i>Dinámica de grupo</i> <i>Gabinets de lectura en bibliotecas</i>	<i>Hornos de microondas</i> <i>Dispositivos ópticos de reconocimiento de caracteres</i>
Término calificado		<i>Radio (metal)</i> <i>Tierra (planeta)</i>	<i>Polarización (luz)</i> <i>Humanidad (teología cristiana)</i>
Frase coordinada		No se acepta	<i>Inflación y contabilidad</i> <i>Reyes y soberanos</i>
Frase múltiple		No se acepta	<i>Escultores españoles</i> <i>[franceses, griegos, etc.]</i>
Frase invertida (sólo como referencia)		No se acepta	<i>Habla, Trastornos del</i> <i>Véase Trastornos del habla</i>

Tabla 2. Formas gramaticales permitidas en tesauros y listas de epígrafes

En ambos casos los términos de indización se basan en la fase nominal, es decir aquélla en la que el núcleo es un sustantivo y el modificador puede ser un adjetivo (modificador directo) o una frase preposicional (modificador indirecto). El plan de revisión de la norma Z39.19 propone reducir el énfasis en este tipo de frase, lo que significaría un cambio sustancial en la concepción de cualquier vocabulario controlado.

Tanto los tesauros como las listas de epígrafes coinciden en que el término de indización simple (una sola palabra) debe ser un sustantivo. En cuanto a los términos compuestos, es decir aquéllos que constituyen una frase debe recordarse que su importancia radica en que mejoran la especificidad de la indización y la precisión de la recuperación. Si como se menciona en el plan de revisión de la norma Z39.19 se intenta sustituir los términos con calificador por términos compuestos, la relevancia que éstos adquieren es aún mayor. En general, la norma ISO 2788 establece que cuando la frase se compone de un núcleo (sustantivo) y un modificador directo (adjetivo) conviene mantener el compuesto. Sin embargo, no se establece claramente un límite para el número de adjetivos permitidos, con el fin de evitar descriptores demasiado largos. Por ejemplo, el concepto vuelos espaciales tripulados podría representarse por dos términos de indización: 1) *Vuelos espaciales*, 2) *Vuelos tripulados*.

Cuando el modificador de la frase es indirecto (frase preposicional) debe considerarse la posibilidad de su fraccionamiento sintáctico. Así el descriptor *Gabinetes de lectura en bibliotecas* podría fraccionarse en tres descriptores *Gabinetes*, *Lectura*, *Bibliotecas*, o en dos descriptores *Gabinetes de lectura*, *Bibliotecas*. Obviamente, debe tenerse cuidado de no fraccionar los términos que puedan perder su significado al romper el compuesto, por ejemplo, *control de calidad*, *mesa redonda*, etc.

Una desventaja del fraccionamiento sintáctico es que aumenta el número de términos tanto en el tesoro como en los registros bibliográficos, pero también es cierto que facilita el uso de los operadores booleanos y la coincidencia entre los términos del usuario y los descriptores. Sería muy difícil que el usuario de un catálogo en línea o de un sitio Web escriba en el buscador un término que coincida exactamente con el descriptor *Evaluación de las tareas inherentes a un empleo*, que es un descriptor del Tesoro de la Unesco.

Contrariamente, si no se fraccionan los descriptores se corre el riesgo de crear inconsistencias en el vocabulario controlado. Por ejemplo, en el Tesoro de la Unesco aparecen los siguientes casos:

- 1) Un descriptor *Enseñanza*, otro descriptor *Geología* y un tercer descriptor *Enseñanza de la Geología*.
- 2) Un descriptor *Enseñanza*, otro descriptor *Medicina* y un tercer descriptor *Enseñanza médica*.
- 3) Un descriptor *Enseñanza* y otro descriptor *Bibliotecología*, pero no un tercer descriptor similar a los mencionados en los casos 1 y 2.

Si se tuviera que indizar un documento sobre la enseñanza de la Bibliotecología, sería necesario asignarle dos descriptores: 1) *Enseñanza*, 2) *Bibliotecología*. Si éste es el criterio, un documento sobre enseñanza de la Geología debería indizarse también con dos descriptores: 2) *Enseñanza*, 2) *Geología*. Sin embargo, existe un descriptor *Enseñanza de la geología* que surge de la aplicación de un criterio diferente evidenciando una inconsistencia del tesoro. Por otro lado, se deja abierta la posibilidad de que en una misma biblioteca un indizador asigne, por ejemplo, el descriptor *Enseñanza de la geología* y otro indizador prefiera asignar dos descriptores 1) *Enseñanza*, 2) *Geología* para representar el mismo concepto, provocando una inconsistencia en el catálogo. Tampoco se entiende por qué en el caso 1 se prefiere una frase preposicional (*Enseñanza de la geología*) y en el caso 2 se prefiere una frase adjetivada (*Enseñanza médica*).

Con respecto a los epígrafes no existen recomendaciones similares para los términos compuestos pero, si se aplicara la norma ISO 2788, un epígrafe como *Dispositivos ópticos de reconocimiento de caracteres* que se encuentra en LEMB, debería fraccionarse en dos epígrafes: 1) *Dispositivos ópticos*, 2) *Reconocimiento de caracteres*, que serían más adecuados para la recuperación. Por otra parte las frases coordinadas como *Inflación y contabilidad*, o *Reyes y soberanos*, no tienen razón de ser en un sistema de información que permita utilizar el operador booleano *and* y sí pueden ser un serio inconveniente para la recuperación.

Todo lo anterior demuestra que el tratamiento de los términos compuestos, en general, ha sido descuidado por los vocabularios controlados a pesar de las reglas explícitas que proporciona la norma ISO 2788 y las recomendaciones de otros vocabularios como las del *Art and architecture thesaurus* que han sido incluidas en el Manual de Aitchison et al. (1997:181-185).

La frase invertida no está permitida por la norma ISO 2788 para ningún idioma. Aunque algunas listas de epígrafes la mantienen, LEMB acertadamente la discontinuó desde su segunda edición de 1985 y sólo la conserva como forma variante, es decir para remitir al epígrafe. De todos modos, deben recordarse dos cosas importantes: a) los vocabularios controlados se rigen, entre otros, por el principio del orden de la lengua natural y, dado que en la lengua española el sustantivo se antepone al adjetivo, no se justifica de ninguna manera la inversión como se podría argumentar para el inglés, b) los sistemas en línea, que permiten buscar una determinada palabra sin importar el orden que ocupe en una frase, hacen innecesaria la frase invertida.

En la Tabla 3 se compara el uso del singular y del plural. Las recomendaciones de la norma ISO 2788 son claras y simples. En el caso de los epígrafes son un tanto dispersas (Rovira, 1966, Westby, 1984) pero, en general, se consideran coincidentes.

Norma ISO 2788	Tesauros	Listas de epígrafes
Entidades concretas cuantificables: en plural	<i>Impresoras</i> <i>Profetas</i>	<i>Cometas</i> <i>Novelistas</i>
Entidades concretas no cuantificables: en singular	<i>Boro</i> <i>Nitrógeno</i>	<i>Quinina</i> <i>Trigo</i>
Entidades abstractas: singular	<i>Cristianismo</i> <i>Hidrografía</i>	<i>Descentralización</i> <i>Escepticismo</i>
Excepción: entidades abstractas que representen una clase con dos o más miembros: plural	<i>Ciencias físicas</i> <i>Reacciones termonucleares</i>	<i>Ciencias políticas</i> <i>Reacciones químicas</i>

Tabla 3. Plural y singular en tesauros y listas de epígrafes

Con respecto a los términos que puedan resultar ambiguos, la norma ISO 2788 ofrece dos maneras de diferenciarlos:

a) Cuando se trata de un homónimo o un término polisémico se agrega un calificador, entre paréntesis, que es parte del descriptor. Este recurso es común a los tesauros y a las listas de epígrafes, como lo demuestran los ejemplos incluidos en la Tabla 2. El plan de revisión de la norma Z39.19 contempla la sustitución de los descriptores con calificador por términos compuestos, hecho que tendría dos consecuencias: 1) una mayor similitud con la lengua natural, ya que el calificador entre paréntesis es un artificio del control de vocabulario y 2) evitar el uso de signos de puntuación –en este caso el paréntesis– que suelen presentar inconvenientes para la recuperación en bases de datos.

b) Cuando se trata de aclarar el significado de un término se puede agregar una nota de alcance. Tanto los tesauros como las listas de epígrafes aceptan varios tipos de notas de alcance: 1) la definición del término; 2) una limitación del alcance del término que puede ser positiva (el término incluye ...) o negativa (el término excluye ...); 3) una instrucción para los indizadores; 4) una nota histórica que indique por ejemplo en qué fecha cambió de nombre un país o una institución; 5) las fechas en que un término es dado de baja, sustituido por otro o subdividido en dos o más.

Es interesante llamar la atención respecto de que algunos autores sugieren agregar una definición a todos y cada uno de los términos de un vocabulario controlado (Aguado et al., 1999; Buchan, 1989; Hudon, 1996; Sager y L'Homme, 1994; Sager y Ndi-Kimbi, 1995; Svenonius, 1997). Hay dos propuestas al respecto: agregar la definición en el vocabulario controlado o vincular el vocabulario controlado con un diccionario. Esta última idea ha tomado fuerza en la medida en que la World Wide Web facilita la vinculación, como sugiere el plan

de revisión de la norma Z39.19. Sin embargo, hay que tener en cuenta dos cosas: 1) un término controlado, ya sea descriptor o epígrafe, sólo permite una definición (de lo contrario no sería controlado) y en un diccionario el mismo término puede tener múltiples acepciones; 2) en el caso de los epígrafes, la precoordinación sería un inconveniente serio para remitir a un diccionario, ya que una cadena precoordinada consta de varios términos y si cada uno de ellos se vincula con múltiples acepciones, podría llegarse a una situación caótica.

Semejanzas y diferencias en el tratamiento de las relaciones

La relación equivalente recibe un tratamiento similar en los tesauros y las listas de epígrafes. Las diferencias tienen que ver con la simbología que indica la relación (Use en lugar de Véase y UP en lugar de x), pero no con la concepción de la relación, según se puede observar en la Tabla 4:

Término	Tesauros	Listas de epígrafes
No descriptor o forma variante	<i>Era terciaria</i> <i>USE Período terciario</i>	<i>Sentencias morales</i> <i>Véase Proverbios</i>
Descriptor o epígrafe	<i>Período terciario</i> <i>UP Era terciaria</i>	<i>Proverbios</i> <i>x Sentencias morales</i>

Tabla 4. La relación de equivalencia en tesauros y listas de epígrafes

En cambio, es notoria la diferencia a favor de los tesauros en cuanto al tratamiento de las relaciones jerárquicas y asociativas. La norma ISO 2788 establece cuatro tipos de relación jerárquica: género/especie, todo/parte, enumerativa y polijerárquica. Ninguna de las listas de epígrafes conocidas en nuestro medio (LCSH, LEMB, Sears), ni las recomendaciones para su redacción (Chan, 1990; Rovira, 1966; Westby, 1984) han contemplado una regla similar y por lo tanto el tratamiento de las relaciones jerárquicas en las listas de epígrafes es muy deficiente.

Una recomendación aplicable a estas listas (Rovira, 1966; Westby, 1984) es que cuando se remite de un término más general a otro más específico se establece una relación de véase además (Va); al contrario cuando se remite de un término más específico a otro más general la relación es una referencia segunda o xx. Debido a que la misma simbología se utiliza para indicar la relación asociativa, en la práctica no hay una cabal distinción entre las relaciones, como se puede observar en el siguiente epígrafe de LEMB:

Mapas
Va Atlas
Cartas aeronáuticas
Cartas náuticas
Cartografía
Escalas (cartografía)
Mapamundis

Mientras que *Atlas*, *Cartas aeronáuticas*, *Cartas náuticas* y *Mapamundis* pueden reconocerse como términos específicos de *Mapas*, no lo son de ninguna manera ni *Cartografía* ni *Escalas (cartografía)*, que son relaciones asociativas. De modo que un primer disparate es utilizar la misma simbología para dos tipos de relaciones diferentes.

Ahora bien, la relación jerárquica vincula dos términos que son epígrafes, pero no en la misma categoría, pues uno está subordinado al otro y viceversa. En este caso es correcto que se utilicen dos símbolos diferente (*Va/xx*), como sucede también en los tesauros (*TE/TG*). Pero la relación asociativa vincula dos términos que tienen la misma categoría en la relación, de modo que deberían tener un solo símbolo como sucede en los tesauros (*TR/TR*). Un segundo disparate, entonces, es asignarle dos símbolos diferentes a un mismo tipo de término en la relación asociativa.

Las incoherencias que se producen por esta situación llegan a tal punto que es posible encontrar en LEMB el siguiente ejemplo:

Bibliotecas de investigación
Va Bibliotecas universitarias
xx Bibliotecas universitarias

Si se asume que esto es una relación jerárquica, ¿cómo es posible que un término sea al mismo tiempo más general y más específico de otro? Si es una relación asociativa, ¿por qué se repite una vez con *Va* y otra con *xx*? ¿La primera relación es asociativa y la segunda jerárquica? No hay una respuesta satisfactoria que sea evidente para el indizador, pero sí infinidad de problemas para el tratamiento de estas relaciones en las bases de datos de las listas de autoridades de materia.

Las listas de epígrafes cuentan con una modalidad adicional que es la referencia general. Se trata en realidad de una relación que puede ser jerárquica o asociativa, pero en vez de remitir a otros términos uno por uno, lo hace en forma general mediante una nota. Por ejemplo:

Aves
Va Animales y aves de caza
Aves acuáticas
Aves de jaula
Aves de rapiña
Y los nombres de determinadas aves, ej. Canarios, Petirrojos, etc.

Esta referencia general resultaba muy útil en las listas impresas para vincular un epígrafe con otros que a veces no se pueden agotar o predecir en una lista, por ejemplo, los escritores, los deportistas, etc. Sin embargo, en los sistemas en línea es deseable que los términos estén vinculados directamente para facilitar la navegación y por lo tanto la referencia general no tiene mucho sentido.

Por otra parte, los vocabularios controlados pueden ser monolingües o multilingües. En este último caso, algunos han optado por indicar la simbología de las relaciones con las abreviaturas en uno de los idiomas, por ejemplo el inglés. Así, en la edición en español del Tesoro de la Unesco se indican las relaciones de esta manera:

Sufragio femenino
UF Sufragistas
BT Derecho de la mujer
Representación política

Es decir *UF* (*used for*) en lugar de *UP* (usado por), *BT* (*broader term*) en lugar de *TG* (término genérico), etc. Como alternativa, tanto las normas ISO 2788 como las GSARE permiten indicar las relaciones con símbolos gráficos en lugar de abreviaturas. En la Tabla 5 se muestran las abreviaturas en español y sus correspondientes símbolos gráficos.

Relaciones	Tesoros	Listas de epígrafes
Nota de alcance	NA	
Equivalencia lingüística (idioma)	X:	=
Equivalencia (sinonimia)	Use/UP, →/=	Véase/x, v/x, →/←-, ->/<-, >/<
Jerarquía	TG/TE, </>	Véase además/xx, va/xx, ->>/<<-, >>/<<
Asociación	TR/TR, -/-	Véase además/xx, va/xx, ->>/<<-, >>/<<
Fraccionamiento sintáctico	+	No se hace

Tabla 5. Abreviaturas y símbolos gráficos utilizados en tesoros y listas de epígrafes

Algunos signos se han usado con distinto significado. Por ejemplo, mientras que en los tesauros los signos > y < indican una relación jerárquica, en las listas de epígrafes indican una relación de equivalencia. El signo = se utiliza en los tesauros en sustitución de UF (relación de equivalencia) y en las listas de epígrafes para introducir el término en otro idioma. No se comprende por qué la norma ISO 2788 recomienda utilizar una sola letra para indicar el idioma de la traducción, cuando se podría utilizar el código ISO de idiomas.

Precoordinación y poscoordinación

En 1950 el bibliotecario estadounidense Mortimer Taube definió su teoría de la indización coordinada (en realidad poscoordinada) que aplicó al sistema *Uniterm* creado por él. Este sistema, aún manual, realizaba la descripción de contenido utilizando fichas (de columnas, de coincidencia óptica o *peek-a-boo*, *zator*, etc.) que se encabezaban con un término único (de allí *uniterm* o unitérmino) y luego se coordinaban en el momento de la recuperación. Interesa aquí rescatar que la teoría de Taube marcó una línea divisoria, a partir de la cual se definieron dos tipos de coordinación: la precoordinación y la poscoordinación

En la *precoordinación* el bibliotecario construye una cadena precoordinada de términos, siguiendo un orden de cita preestablecido y lo asigna al registro bibliográfico antes de su almacenamiento en el catálogo o en la bibliografía. En la mayoría de las listas de epígrafes como LCSH o LEMB, la cadena precoordinada se construye con cinco posibles componentes: un epígrafe principal y cuatro subdivisiones comunes de asunto, espacio, tiempo y forma, que se coordinan exactamente en ese orden (Tabla 6). El epígrafe principal sólo puede ser uno, pero las subdivisiones pueden repetirse, es decir puede haber dos o más subdivisiones de asunto, o dos o más de espacio, etc. Obviamente mientras más larga sea la cadena precoordinada más problemas presentará para su recuperación en línea.

Epígrafe principal	Subdivisiones			
	Asunto	Espacio	Tiempo	Forma
<i>Geología</i>	<i>Enseñanza</i>	<i>Argentina</i>	<i>1990-1999</i>	<i>Informes</i>
No repetible	Repetible	Repetible	Repetible	Repetible

Tabla 6. Matriz para construir la cadena precoordinada utilizando LEMB

La precoordinación tuvo su origen en la necesidad de ofrecer la mayor cantidad de información posible en una única ficha de 7,5 x 12,5 cm. Al introducirse la computadora la precoordinación fue vista como una desventaja

y a esto se debe, entre otras cosas, el éxito de los tesauros. Sin embargo, en la actualidad se le reconocen ciertas ventajas a los sistemas precoordinaados.

Numerosos estudios han demostrado un marcado rechazo de los usuarios de catálogos en línea a utilizar los operadores booleanos, con lo cual se desaprovecha el recurso tal vez más potente de la computadora (Ensor, 1992; Fidel, 1991; Millsap y Ferl, 1993; Olsen et al., 1989; Peters, 1989). Para subsanar este problema, se ha sugerido que los sistemas de información le muestren al usuario una lista de términos -por ejemplo el índice invertido de una base de datos- de la que puedan seleccionar y marcar el término de búsqueda. Es aquí donde la cadena precoordinaada brinda su mayor ventaja.

Los formatos tipo MARC destinan varios subcampos para el registro de la cadena precoordinaada. En el MARC21st esos subcampos son: \$a Epígrafe principal, \$x Subdivisión de asunto, \$y Subdivisión geográfica, \$z Subdivisión cronológica y \$v Subdivisión de forma. Si se necesita indizar un informe sobre la enseñanza de la Geología en Argentina durante la década de los noventa, usando LEMB como vocabulario controlado (Tabla 6) y el formato de descripción MARC21st, se tendría: \$aGeología\$xEnseñanza\$yArgentina\$z1990-1999\$vInformes. La visualización del índice invertido podría mostrar lo siguiente:

1990-1999
Argentina
Enseñanza
Geología
Geología-Enseñanza-Argentina-1990-1999-
Informes

La visualización de la cadena precoordinaada le ofrece al usuario mayor información y limita claramente el alcance del tema tratado en el documento, produciendo una recuperación más precisa. Este es un recurso que no se logra con una indización poscoordinaada convencional. Además, si el usuario de este índice invertido marca el término *Geología*, obtendría todos los registros en los que se ha asignado ese epígrafe. Al mismo tiempo, como cada componente de la cadena precoordinaada está registrado en un subcampo diferente, es posible poscoordinar los términos con los operadores booleanos, por ejemplo *Geología* AND *Argentina*. De modo que estos sistemas ofrecen grandes ventajas, siempre y cuando se registren en subcampos diferentes como se indicó más arriba, pues armonizan la precoordinaación y la poscoordinación para brindarle al usuario más opciones y, por ende, más oportunidades de lograr una búsqueda exitosa.

Las listas de epígrafes que se rigen por los principios o la práctica de LCSH, entre ellas las más conocidas listas de epígrafes en español, presentan serias inconsistencias en la construcción de la cadena precoordinaada, que se arrastran desde los tiempos de Charles A. Cutter. La más conocida es la que se produce al tener que decidir si la materia se subdivide por lugar o si el lugar se

subdivide por materia (Foskett, 1996). Por el contrario, entre 1960 y 1990 se desarrollaron en Gran Bretaña varios sistemas que supieron conjugar la precoordinación y la poscoordinación en sistemas computarizados, pero basados en el método del análisis por faceta inventado por Ranganathan, lo que les dio un sólido fundamento teórico y les permitió eliminar las inconsistencias. Los dos ejemplos más destacados fueron el sistema de indización del *British technology index* ideado por Eric Coates (1988) y el famoso sistema PRECIS (Austin, 1974), considerado como el mejor sistema de indización desarrollado hasta el presente (Dykstra, 1989) y que lamentablemente debió ser abandonado en aras de la cooperación entre British Library y Library of Congress.

En 1998 OCLC decidió optimizar LCSH para adaptarla mejor a los registros metadatos. La iniciativa, conocida como *Faceted application of subject terminology* (FAST) propone mantener el vocabulario de esta lista que en la actualidad cuenta con más de 200.000 epígrafes, pero simplificando su sintaxis, ya que los términos se organizan en un sistema facetado y computarizado (O'Neill y Chan, 2003). Lamentablemente el sistema FAST se basa en consideraciones eminentemente prácticas que lo colocan muy lejos de sus antecesores británicos.

Conclusiones

En primer lugar, consideramos que la propuesta de revisar la norma estadounidense Z39.19 para adaptarla a las nuevas tecnologías es una excelente iniciativa que debería ser imitada por otros organismos normalizadores, especialmente de alcance internacional. Es de resaltar además la intención de extender esta normativa a las listas de epígrafes, que han carecido de una norma coherente durante sus más de cien años de existencia. Sin embargo, hay ciertos aspectos del plan de revisión que deben llamar nuestra atención.

En cuanto al vocabulario en sí y a sus relaciones semánticas, consideramos que es factible y deseable establecer normas comunes para ambos tipos de SOC. No obstante, nos preocupa la intención de hacer las formas gramaticales menos rígidas, restarle énfasis a la frase nominal y vincular los términos con un diccionario, así como minimizar la importancia de las relaciones. El control de vocabulario debe ser riguroso para compensar las dificultades que ofrece la lengua natural en la recuperación (sinonimia, homonimia, polisemia, etc.) y en este sentido creemos conveniente recordar que el vocabulario controlado y la lengua natural producen recuperaciones distintas, aunque complementarias (Dubois, 1987; Svenonius, 1986):

- La lengua natural (cuando se recupera sobre un resumen o el texto completo) aumenta la precisión por el nivel de especificidad y la actualización de los términos, mientras que el vocabulario controlado aumenta la precisión mediante el uso de los términos compuestos y el control de homónimos y térmi-

nos polisémicos.

- La lengua natural aumenta la tasa de acierto por la mayor cantidad de términos incluidos en un resumen o un texto completo, mientras que el vocabulario controlado aumenta la tasa de acierto mediante las relaciones equivalentes, jerárquicas y asociativas.

Si este control se hace menos rígido, ¿qué sentido tiene un vocabulario controlado? Sólo en la medida en que se combinen los dos tipos de lenguas en un mismo sistema de recuperación de información se amplían las oportunidades de que el usuario logre una búsqueda exitosa.

Con respecto a la coordinación, creemos conveniente que antes de invertir tiempo, dinero y esfuerzo en normalizar las subdivisiones de las listas de epígrafes, se promueva la poscoordinación y se investigue a fondo la posibilidad de gestionar los términos controlados poscoordinados en un sistema facetado y computarizado, rescatando el modelo de PRECIS y optimizándolo por medio de las nuevas tecnologías.

En nuestra opinión, la única forma de integrar las listas de epígrafes con los tesauros, es convertir a las primeras en SOC poscoordinados, regidos por la norma ISO 2788 o equivalente. Es importante destacar que las listas de epígrafes existentes en todo el mundo contienen un vocabulario riquísimo que puede ser la base de una lista de autoridades de materia universal en cuanto a la temática y multilingüe pues se han desarrollado en diferentes idiomas; desde este punto de vista son absolutamente rescatables. Sin embargo, la necesidad de mejorar la definición de las relaciones jerárquicas y asociativas, así como las dificultades de la coordinación, pueden resultar en un emprendimiento demasiado costoso.

Como quiera que sea, debemos esperar cambios importantes en el diseño y la gestión de los vocabularios controlados tradicionales.

Nota

¹ *Sistemas de organización de conocimiento* es un nuevo término para referirse a lo que antes se denominaba *lenguajes documentales* (Hodge, 2000).

Referencias bibliográficas

Aguado, Amelia; Rosa Z. Pisarello; Ana M. Martínez y Cecilia Delorenci. 1999. Definición: nuevas tendencias en tesauros y otros vocabularios controlados. En: López Huertas, María José y Juan Carlos Fernández Molina, editores. La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información: Actas del Congreso ISKO-España EOCONSID'99 (4: 1999: Granada). Granada: Universidad de Granada, p. 127-132.

- Aitchison, Jean. 1995. *Tesaurus de la Unesco*. 3ª ed. París: Unesco. 716 p.
- Aitchison, Jean; Alan Gilchrist y David Bawden. 1997. *Thesaurus construction*. 3ª ed. London: Aslib. 211 p.
- Asociación Española de Normalización. 1989. Documentación: directrices para el establecimiento y desarrollo de tesauros monolingües. Parte I. En *Revista española de documentación científica*. Vol. 12, no. 4, 463-483.
- Asociación Española de Normalización. 1990. Documentación: directrices para el establecimiento y desarrollo de tesauros monolingües. Parte II. En *Revista española de documentación científica*. Vol. 13, no. 1, 601-629.
- Austin, Derek. 1974. *PRECIS: a manual of concept analysis and subject indexing*. London: Council of the British National Bibliography. 551 p.
- Bates, Marcia. 1988. How to use controlled vocabularies more effectively in online searching. En *Online*. Vol. 12, no. 6, 45-56.
- Biblioteca Luis Angel Arango. 1998. *Lista de encabezamientos de materia para bibliotecas*. 3ª ed. Bogotá: Rojas Eberhard, 2 v.
- Bloomfield, Masse. 1993. A look at subject headings: a plea for standardization. En *Cataloging and classification quarterly*. Vol. 16, no. 1, 199-124.
- Buchan, Ronald L. 1989. Intertwining thesauri and dictionaries. En *Information services and use*. Vol. 9, 171-175.
- Chan, Lois Mai. 1989. A subject cataloging code? En *Cataloging and classification quarterly*. Vol. 10, no. 2, 199-202.
- Chan, Lois Mai. 1990. *Library of Congress subject headings: principles of structure and policies for application*. Washington, DC: Library of Congress. <<http://www.tlcdelivers.com/tlc/crs/shed0014.htm>> [Consulta: 8 octubre 2003].
- Coates, Eric. 1988. *Subject catalogues: headings and structures*. 2ª ed. London: Library Association Publishing. 186 p.
- Cochrane, Pauline A. 1984. Modern subject access in the online age. En *American libraries*. Vol. 15, no. 2, 80-83; no. 3, 145-148; no. 4, 250-252; no. 5, 336-339; no. 6, 438-441; no. 7, 527-529.
- Dubois, Pablo. 1987. Free text vs controlled vocabulary: a reassessment. En *Online review*. Vol. 11, no. 4, 243-253.
- Dykstra, Mary. 1988. LC subject headings disguised as a thesaurus. En *Library journal*. Vol. 113, no. 1, 42-46.

- Dykstra, Mary. 1989. PRECIS in the online catalog. En *Cataloging and classification quarterly*. Vol. 10, no. 2, 81-94.
- Ensor, Pat. 1992. User practices in keyword and Boolean searching on the online public access catalog. En *Information technology and libraries*. Vol. 11, no. 3, 210-219.
- Fidel, Raya. 1992. Who needs controlled vocabulary? En *Special libraries*. Vol. 83, no. 1, 1-9.
- Foskett, Antony C. 1996. The subject approach to information. 5ª ed. London: Library Association Publishing. 456 p.
- Gilchrist, Alan y Peter Kibby. 2000. Taxonomies for business: access and connectivity in a wired world. London: TFPL Ltd. 175 p.
- Henige, David. 1987. Library of Congress subject headings: is euthanasia the answer? En *Cataloging and classification quarterly*. Vol. 8, no. 1, 7-19, 21-23.
- Hodge, Gail. 2000. Systems of knowledge organization for digital libraries: beyond traditional authority files. S.L.: Council on Library and Information Resources; Digital Library Federation. <<http://www.clir.org/pubs/reports/pub91/contents.html>> [Consulta: 9 septiembre 2003].
- Holley, Robert P. y Robert E. Killheffer. 1982. Is there an answer to the subject access crisis? En *Cataloging and classification quarterly*. Vol. 1, no. 2/3, 125-133.
- Hudon, Michele. 1996. Preparing terminological definitions for indexing and retrieval thesauri: a model. En: Knowledge organization and change: Proceedings of the 4th International ISKO Conference, Washington DC, 15-18 July 1996. En *Advances in Knowledge Organization*. Vol. 5, 363-369.
- Instituto Argentino de Normalización. 1983. Tesoros monolingües para la recuperación de la información: estructura y desarrollo. Buenos Aires: El Instituto (Norma IRAM 32057). 35 p.
- International Federation of Library Associations and Institutions. 1995. Directrices para los registros de autoridad y referencia de materia. Madrid: ANABAD. 78 p.
- International Organization for Standardization. 1986. Documentation: Guidelines for the establishment and development of monolingual thesauri. 2nd ed. Geneva: The Organization (ISO 2788). 32 p.

- Klugman, Simone. 1989. Failures in subject retrieval. En *Cataloging and classification quarterly*. Vol. 10, no. 2, 9-35.
- Lancaster, Frederick W. 1995. El control del vocabulario en la recuperación de información. Valencia: Universitat de Valencia. (Educació: Materials 12). 286 p.
- Martínez, Ana M.; Cecilia Delorenzi; Rosa Z. Pisarello y Amelia Aguado. 2003. Variantes en el tratamiento de términos prestados en la terminología bibliotecológica y documental en español: los lenguajes documentales. En *Revista Argentina de Bibliotecología*. Vol. 3 [aceptado para su publicación].
- Millsap, Larry y Terry E. Ferl. 1993. Search patterns of remote users: an analysis of OPAC transaction logs. En *Information technology and libraries*. Vol. 12, no. 3, 321-343.
- National Information Standards Organization. 1993. Guidelines for the construction, format, and management of monolingual thesauri. Bethesda: The Organization (ANSI/NISO Z39.19). 69 p.
- National Information Standards Organization. 2003. Developing the next generation of standards for controlled vocabularies and thesauri. Bethesda: The Organization. <<http://www.niso.org/committees/MT-info.html>> [Consulta: 3 junio 2003].
- Olsen, Randy J.; John O. Christensen; Kal A. Larsen y Kayla Willey. 1989. Implementing NOTIS keyword/Boolean searching: a case study. En *Cataloging and classification quarterly*. Vol. 10, no. 2, 125-140.
- O'Neill, Edward T. y Lois Mai Chan. 2003. FAST (faceted application of subject terminology): a simplified LCSH-based vocabulary. En IFLA General Conference and Council (69th: 2003: Berlín). [S.l.]: IFLA. <http://www.ifla.org/IV/ifla69/papers/010e-Oneill_Mai-Chan.pdf> [Consulta: 29 agosto 2003].
- Peters, Thomas A. 1989. When smart people fail: an analysis of the transaction log of an online public access catalog. En *Journal of academic librarianship*. Vol. 15, no. 5, 267-273.
- Reynolds, Sally J. 1989. In theory there is no solution: the impediments to a subject cataloging code. En *Library quarterly*. Vol. 59, no. 3, 223-238.
- Rolland-Thomas, Paule. 1993. Thesaural codes: an appraisal of their use in the Library of Congress subject headings. En *Cataloging and classification quarterly*. Vol. 16, no. 2, 71-91.

- Rovira, Carmen. 1966. Los epígrafes en el catálogo diccionario. 2ª ed. Washington DC: Unión Panamericana. 134 p.
- Sager, Juan C. y Marie Claude L'Homme. 1994. A model for the definition of concept rules for analytical definitions in terminological databases. En *Terminology*. Vol. 1, 351-353.
- Sager, Juan C. y Augustin Ndi-Kimbi. 1995. The conceptual structure of terminological definitions and their realizations: a report on research in progress. En *Terminology*. Vol. 2, 61-81.
- Shubert, Steven B. 1992. Critical views of LCSH ten years later: a bibliographic essay. En *Cataloging and classification quarterly*. Vol. 15, no. 1, 37-97.
- Studwell, William E. 1993. LC's head in the sand, or, why the subject cataloging manual is not enough. En *Technical services quarterly*. Vol. 10, no. 3, 45-50.
- Studdwell, William E. y Steve L. Folsom. 1993. Game shows, elevators, full plates, and other allegories: a look at the present state and future possibilities of LC subject headings. *Serials librarian*. Vol. 23, no. 3/4, 241-244.
- Svenonius, Elaine. 1986. Unanswered questions in the design of controlled vocabularies. En *Journal of the American Society for Information Science*. Vol. 37, no. 5, 331-340.
- Svenonius, Elaine. 1997. Definitional approaches in the design of classification and thesauri and their implications for retrieval and for automatic classification. En: Knowledge organization for information retrieval: Proceedings of the International Study Conference on Classification Research, University College (6: 1997: London). Hague: International Federation of Information and Documentation. p. 12-16.
- Westby, Barbara M. 1984. Principios que rigen la lista de encabezamientos de materia de Sears. En Westby, Barbara M. Sears: lista de encabezamientos de materia. 2ª ed. Nueva York: H.W. Wilson. p. xiii-xlv.