

Control de vocabulario: orígenes, evolución y proyección

Mario Barité

Doutor em Ciência da Informação pela Universidad de Granada – Granada, Espanha. Professor da Universidad de la República de Uruguay, Montevideo - Uruguay.

http://buscadores.anii.org.uy/buscador_sni/exportador/ExportarPdf?hash=2e17143bf99bf82b9ec248297b244d01
E-mail: mabarite@gmail.com

Recebido em: 10/07/2014. Aprovado em: 22/06/2015. Publicado em: 15/01/2016.

Resumen

Las cuestiones relacionadas con el control de vocabulario están fuertemente asentadas en la literatura de la Organización del Conocimiento desde los orígenes mismos de este campo temático. El control de vocabulario puede ser visto como un conjunto de técnicas y procedimientos que se aplican sobre el lenguaje para resolver problemas de comprensión, ambigüedad, alcance y relación entre términos que expresan conceptos, y/o entre denominaciones que expresan nombres de personas, lugares, productos o instituciones. Pretende superar las dificultades de interpretación y comunicación propias de la utilización del lenguaje natural, las que son vistas como desviaciones del discurso metódico y preciso que requieren las ciencias y las disciplinas. En este trabajo se recoge la historia del término ‘control de vocabulario’, desde su origen en el ámbito de la enseñanza del inglés como una segunda lengua, a comienzos de la década de 1930, y su pasaje al vocabulario de la Bibliotecología y Ciencia de la Información hacia 1950. Se presentan en forma sistematizada cinco funciones del control de vocabulario: restricción, desambiguación, normalización, jerarquización y relación. A continuación se identifican y explican los nueve procedimientos de control de vocabulario más utilizados. Se analizan elementos del debate histórico entre quienes defienden la utilización de vocabularios controlados y quienes proponen el lenguaje natural como mejores herramientas para la representación temática del contenido de documentos y recursos de información, y se propugna la superación de ese debate en virtud de las características actuales del entorno digital en que se desenvuelve buena parte de los problemas de almacenamiento, organización y recuperación de documentos y datos. Se considera la idea de que la interoperabilidad puede ser vista también como una modalidad particular de control de vocabulario. Se establecen conclusiones respecto a la vigencia y proyección del control de vocabulario.

Palabras clave: Control de vocabulario. Vocabularios controlados. Representación del conocimiento.

Vocabulary control: origins, evolution and future

Abstract

The issues related to vocabulary control are firmly rooted in the literature of Knowledge Organization since the very beginning of this subject area. Vocabulary control may be regarded as a set of techniques and procedures applied to language in order to solve comprehension, ambiguity, scope and relation problems between terms expressing concepts, and/or between designations naming people, places, products or institutions. It intends to overcome interpretation and communication difficulties arising from the use of natural language, which in turn are seen as deviations from the methodical and precise speech required by science and disciplines. This paper records the history of the term “vocabulary control” since its origin in the field of the teaching of English as a second language at the beginning of the 1930 decade, to its insertion in the lexis of Library and Information Science towards 1950. There follow the systematized presentation of five functions of vocabulary control, namely: restriction, disambiguation, standardization, hierarchy and relation, as well as the identification and explanation of the nine most usually employed vocabulary control procedures. It continues with the analysis of some of the elements in discussion in the historical debate between those in favor of the usage of controlled vocabularies and those supporting natural language as a more appropriate tool for the subject representation of the contents of documents and information resources, and the superseding of said debate is ensuingly supported by virtue of the present-day characteristics of the digital means in which a substantial part of the issues bearing on the storage, organization and retrieval of documents and data are dealt with. Finally, the fact that interoperability may be also seen as a particular means of vocabulary control is taken into consideration. Conclusions are provided with regard to effectiveness and prospect of vocabulary control.

Keywords: Vocabulary control. Controlled vocabularies. Knowledge representation.

Controle de vocabulário: origens, evolução e projeção

Resumo

As questões relacionadas com o controle de vocabulário estão fortemente assentadas na literatura da Organização do Conhecimento desde as próprias origens deste campo temático. O controle de vocabulário pode ser visto como um conjunto de técnicas e procedimentos aplicados sobre a linguagem para resolver problemas de compreensão, ambiguidade, alcance e relação entre termos que expressam conceitos, e/ou entre denominações que expressam nomes de pessoas, lugares, produtos ou instituições. Pretende superar as dificuldades de interpretação e comunicação próprias do uso da linguagem natural, as vistas como desvios do discurso metódico e preciso que requerem as ciências e as disciplinas. O trabalho coleta o histórico do termo 'controle de vocabulário', desde sua origem no âmbito do ensino do inglês como um segundo idioma, no início da década de 1930, e sua passagem ao vocabulário da Biblioteconomia e Ciência da Informação perto de 1950. Se apresentados em forma sistematizada cinco funções do controle de vocabulário: restrição, desambiguação, normalização, hierarquização e relação. A seguir são identificados e explicados os nove procedimentos de controle de vocabulário mais utilizados. São analisados elementos do debate histórico entre os que defendem o uso de vocabulários controlados e os que propõem a linguagem natural como melhores ferramentas para a representação temática do conteúdo de documentos e recursos de informação, e se defende a superação desse debate em virtude das características atuais do ambiente digital no qual se desenvolve boa parte dos problemas de armazenamento, organização e recuperação de documentos e dados. A interoperabilidade como uma modalidade particular de controle de vocabulário também é considerada. As conclusões dizem respeito à vigência e projeção do controle de vocabulário.

Palavras-chave: Controle de vocabulário. Vocabulários controlados. Representação do conhecimento.

INTRODUCCIÓN

Con este trabajo se pretende introducir una reconstrucción histórica, conceptual y de método, de la evolución de los procesos de control de vocabulario en relación con los sistemas de organización del conocimiento y con la representación y recuperación temática en sistemas de información.

En primer lugar se pretende revisar y actualizar el concepto de control de vocabulario, ubicando sus procedimientos como formas de intervención desde una perspectiva lingüística.

En segundo lugar, se aspira a ajustar la “historia oficial” de la introducción del término *vocabulary control* en la Organización del Conocimiento, pues tradicionalmente se ha asociado el surgimiento del término con la creación de los primeros tesauros.

En tercer término, se propone una sistematización de las funciones que persigue el control de vocabulario, llevada a cabo a partir del análisis comparado de los principales estándares internacionales en la materia.

A continuación se analiza una enumeración taxativa de los nueve procedimientos de control

de vocabulario que se utilizan habitualmente por parte de responsables del desarrollo de vocabularios controlados y de sistemas de información.

Se estudia luego la situación de los procesos de control de vocabulario en un entorno digital que ha mudado en los últimos veinte años con una dinámica extraordinaria.

Finalmente se establecen conclusiones en relación con los aspectos estudiados.

CONCEPTO DE CONTROL DE VOCABULARIO

Las cuestiones relacionadas con el control de vocabulario están fuertemente asentadas en la literatura de la Organización del Conocimiento desde los orígenes mismos de este campo temático, y han acompañado -a veces lateralmente, a veces con un fuerte protagonismo- el debate de las ideas en torno a las mejores soluciones para representar el contenido de los documentos, y establecer por este medio, vías fiables, rápidas y pertinentes para la búsqueda temática de información.

Sin perjuicio de reconocer el antecedente más antiguo en Konrad Gesner (quien ya en 1548 había establecido en su *Pandectarum* la referencia de véase, así como un sistema de clasificación del conocimiento de veintiún clases), fueron Dewey a través del desarrollo de su sistema de clasificación, y Cutter con la proposición de criterios normalizadores para la presentación de datos en los catálogos, quienes colocaron las piedras fundacionales del edificio donde gobiernan con sus respectivos pabellones la Organización del Conocimiento y la Recuperación de Información.

Si bien la noción de control de vocabulario se introdujo en el área posiblemente a fines de la década de 1940, como se verá, un cuarto de siglo después Lancaster todavía encontraba importante preguntarse *¿why control vocabulary?*, en el primer capítulo de una de sus obras clásicas (LANCASTER, 1972). Es una pregunta que aparece recurrentemente en textos y comunicaciones a congresos, como si aún se necesitara una respuesta concluyente y final. O quizás, en tanto la pregunta ha mantenido su fuente de legitimidad inicial, su respuesta ha ido cambiando con el tiempo, como consecuencia del dinamismo de los fenómenos de la representación del conocimiento.

En una primera aproximación, el control de vocabulario puede ser visto como un conjunto de técnicas y procedimientos que se aplican sobre el lenguaje para resolver problemas de comprensión, ambigüedad, alcance y relación entre términos que expresan conceptos, y/o entre denominaciones que expresan nombres de personas, lugares, productos o instituciones. Su aplicación se fundamenta en la idea de que el lenguaje natural, tal como se expresa en textos y documentos de la más variada especie genera -inclusive en comunidades de usuarios especializadas y más o menos sólidas y cerradas- dificultades de interpretación y de comunicación, las que son vistas como desviaciones del discurso metódico y preciso que requieren las ciencias y las disciplinas en general para su consolidación y avance.

La presumida eficacia de la indización a través de vocabularios controlados por sobre la obtenida a través de la selección directa de expresiones del lenguaje natural ha estado siempre en controversia. El debate entre quienes propugnan que el lenguaje natural se basta para la representación adecuada del conocimiento expuesto en documentos y su recuperación, y quienes sostienen que la vía más eficaz para esos fines se da a través de sistemas desarrollados mediante operaciones de control de vocabulario, es de larga data y no parece tener un punto de definición o equilibrio (AUSTIN, 1986; BATES, 1999; FUCHS, 2010). A menos de doce años de la creación del primer tesoro, ya había quien afirmaba que los vocabularios controlados habían quedado anticuados para la indización, y que el lenguaje natural de la prosa científica era completamente adecuado para ese fin (KLINGBIEL, 1970). Siguen siendo comunes en la literatura los estudios de indización que contrastan los resultados de la aplicación de vocabularios controlados o lenguaje natural (RAO, 1998; MÉNARD, 2009). La generalización de aplicaciones informáticas que optimizan el acceso a información mediante claves lingüísticas no necesariamente controladas han dado nueva dimensión a esta cuestión controversial.

El control de vocabulario implica una intervención desde la perspectiva lingüística. La intervención que se realiza debe ser “inteligente” en varias direcciones: por una parte, exige el conocimiento acabado de las dificultades que puede plantear el tránsito e intercambio de ideas, tal como se expresa y procesa en textos y documentos en uno o varios idiomas, que hoy día se encuentran disponibles, se utilizan y se reusan en diferentes formatos y soportes. En segundo lugar, debe aprovechar las facilidades que ofrecen las más novedosas tecnologías y aplicaciones para la búsqueda y la recuperación de datos e información. Por último, debe establecer mecanismos de organización y normalización que sean consistentes, sustentables y adaptables a la evolución del conocimiento, del lenguaje y de las tecnologías.

En Organización del Conocimiento la formulación de las técnicas y procedimientos de control de vocabulario se asocia históricamente con el desarrollo de los primeros tesauros, que requerían métodos probados y normalizados de formalización de lenguaje técnico y científico, para fines de indización y de recuperación de información (ENGINEERS JOINT COUNCIL, 1965; HYSLOP, 1965; DODEBEI, 2002). En la actualidad, el control de vocabulario es utilizado como herramienta habitual en el proceso de creación, desarrollo o revisión, no ya solo de tesauros, sino de cualquier tipo de sistema de organización del conocimiento, pero también como sostén para la organización formal de contenidos etiquetados en Internet.

Como se introdujo más arriba, el control de vocabulario permite establecer formas normalizadas de: a) términos que expresan conceptos, y que se constituyen en descriptores, encabezamientos de materia o términos de indización del repertorio terminológico respectivo (tesauros, taxonomías, listas de encabezamientos, sistemas de clasificación, etc., pero también diccionarios, glosarios, tesauroglosarios y tipos similares); y, b) denominaciones que expresan nombres de personas, lugares, productos o instituciones, y que se transforman en términos de autoridad dentro de alguno de los repertorios mencionados o en sistemas de información para los cuales esas denominaciones constituyen claves de acceso temático.

Del mismo modo, promueve la representación de relaciones entre conceptos, las cuales van reconstruyendo por acumulación y continuidad, la estructura conceptual de un dominio especializado.

Desde una perspectiva más amplia, el control de vocabulario no solo participa de la resolución de los problemas ya mencionados (comprensión, ambigüedad, alcance, relación, interpretación y comunicación), sino que agrega al menos dos cualidades de importancia sustantiva. La primera tiene que ver con su capacidad para mejorar la eficacia y la eficiencia de sistemas de información

(NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010, p. 1), pero también de sistemas de navegación web y de otros sistemas que se construyen a partir de la utilización del lenguaje para describir la realidad. En esta línea el principio de consistencia, entendido como la “coherencia en la aplicación de principios, reglas y procedimientos en lo relativo a la creación y uso de estructuras conceptuales o lenguaje natural para fines de clasificación o indización” (BARITÉ et al., 2013), guarda una relación directamente proporcional con el control de vocabulario, pues se presume que a mayor control de vocabulario, mayor consistencia, y por ende, mayor eficiencia del sistema de información, en cuanto logra facilitar la descripción temática, la búsqueda y la recuperación.

La segunda es su aptitud para organizar campos conceptuales, áreas temáticas y disciplinas a través de la destilación de los conceptos mayormente asociados entre sí en la literatura, sea que estén claramente establecidos y aceptados, o sea que se encuentren en cuestión. De este modo, los vocabularios controlados ofrecen una placa fija del estado del arte de una especialidad en un momento determinado y desde una determinada perspectiva, considerando tanto las prácticas, las teorías y los tópicos ya consolidados como sus puntos de ruptura y conflicto. En su evolución esa placa se transforma en móvil, exhibiendo los cambios producidos dentro de un campo temático a lo largo del tiempo.

Una revisión de las cuestiones relacionadas con el control de vocabulario se justifica en que, por una parte, este tópico se sigue considerando como nuclear en la formación universitaria en Bibliotecología y Ciencia de la Información, tanto a nivel de grado como de posgrado, como evidencia cualquier análisis sumario que se puede realizar sobre programas de estudio y *syllabi* de universidades de todos los continentes. La vigencia del control de vocabulario se expresa también en artículos de revistas y comunicaciones a congresos en todo el mundo, que regularmente abordan algunos de sus aspectos, o miden sus resultados

o su efectividad en sistemas de información. La publicación y actualización de estándares orientados a establecer lineamientos y criterios para el control de vocabulario confirma la vigencia de la temática, así como la necesidad de estudiar críticamente su evolución y proyección, considerando sobre todo el presunto antagonismo entre vocabularios controlados y lenguaje natural.

LOS ORÍGENES

Se carece de una respuesta precisa respecto al momento a partir del cual se comienza a utilizar el término “control de vocabulario” (en inglés *vocabulary control*) en la literatura bibliotecológica, pero al menos es posible reconstruir gran parte de su derrotero.

Como en muchas otras historias de términos, la práctica precedió a la acuñación, y fue necesario reconocer la identidad y la relevancia de esa práctica para asignarle una denominación adecuada. Charles Ammi Cutter ya había presentado algunas reglas para la construcción de encabezamientos de materia alfabéticos en una de sus obras magnas: *Rules for a Dictionary Catalogue*, aparecida en 1876. El segundo hito en la materia lo constituyó la publicación, en 1895, de la *American Library Association's List of Subject Headings for Use in Dictionary Catalogs*, que fue calificada posteriormente como “el primer vocabulario controlado para catálogos alfabéticos por materia” (LANCASTER, 1986, p. 31). Un poco más adelante, entre los años 1909 y 1914, fue publicada por entregas la que es considerada primera edición de la lista de encabezamientos de materia de la Biblioteca del Congreso, que llevó el nombre de *Subject Headings Used in the Dictionary Catalogues of the Library of Congress* (STONE, 2000).

Sin embargo, debieron pasar más de cuarenta años de una práctica regular y cada vez más consolidada de procedimientos ordenados de creación de encabezamientos de materia, para que comenzara a generalizarse la expresión *vocabulary control*. La historia de la acuñación y evolución de este término es, como se verá, un notable ejemplo de dos procesos habituales en el mundo del saber: i) el tránsito de un

término desde un área temática a otra, trasladando la esencia del concepto e impregnándola, al mismo tiempo, con atributos y matices nuevos, propios del área que lo recibe; ii) el préstamo del término se da una vez consolidadas un conjunto de prácticas en el área de recepción, que estaban requiriendo una denominación que les diera una identidad propia, luego de un período en que los especialistas no logran dar a esas prácticas más que una terminología indecisa, y por consecuencia, provisional.

Es posible asegurar que el término *vocabulary control* fue extrapolado desde el área del aprendizaje y el estudio de las lenguas extranjeras hacia la Bibliotecología y la Ciencia de la Información. En efecto, en 1930 el lingüista y filósofo Charles Kay Ogden, publicó un corpus de vocabulario del inglés que incluía unas 850 palabras, al que dio en llamar *Basic English* (OGDEN, 1930). Con este corpus pretendía favorecer la enseñanza de ese idioma como segunda lengua, reduciendo al mínimo el vocabulario necesario para comunicarse. La obra de Ogden, luego asociado con I.A. Richards, fue criticada por considerarse que la lista era insuficiente, no siempre pertinente, y sin el sustento de una visión de conjunto de la lengua.

En parte como reacción al *Basic English* surgió hacia 1930 el Movimiento del Vocabulario Controlado (*Vocabulary Control Movement*), el cual de la mano de Harold Palmer y Michael West generó alternativas fundamentales y precursoras de los diccionarios de aprendizaje de una lengua y de las listas de palabras altamente estructuradas (PALMER, 1938; WEST, 1953; COWIE, 1999). Palmer utiliza inclusive la expresión *vocabulary control* en el título de uno de sus trabajos publicado en dos entregas (PALMER, 1936a; PALMER, 1936b). Este movimiento aportó un nuevo método que “intentó buscar criterios válidos para la selección de vocabulario con la finalidad de simplificar los textos utilizados en la enseñanza de la lectura en lengua extranjera” (ESTEBAN GARCÍA, 2007, p. 172). Uno de esos criterios, el de dar primacía a la dimensión cuantitativa de la aparición de las

palabras en los textos de educación básica, creando para ese fin listas de frecuencias, fue establecido por Thorndike y Lorge (1944), y dio lugar a derivaciones e impactos en diversas áreas del conocimiento tales como la Neuropsicología, la Lingüística, la Lexicografía, la Psicología y la Pedagogía (VARELA BARRAZA et al., 2013). Sería ocioso enfatizar en la importancia que ha tenido históricamente la consideración de la frecuencia de aparición de palabras, en experiencias de indización y recuperación automática de información, y de -incluso- derivaciones tan pintorescas como la creación automática de nubes de palabras a partir del barrido de textos (v. por ejemplo, <http://www.wordle.net>).

Los autores involucrados con el *Vocabulary Control Movement* desarrollaron teoría y métodos en torno, por ejemplo, a la forma de utilizar vocabularios controlados de inglés básico considerando la edad, la madurez psicológica u otras características de las personas dedicadas al aprendizaje de una lengua no nativa, incorporando de este modo una implícita perspectiva de usuario. Por otra parte, algunos de los principios de control de vocabulario propuestos, como la división en categorías conceptuales de las palabras objeto de aprendizaje (concreto, abstracto, estructural) y la organización de expresiones en series sucesivas de *plateaux* verbales (MORRIS, 1947) preanuncian procedimientos y elementos metodológicos que serán replicados y adaptados más adelante, para el desarrollo de tesauros y otros vocabularios.

En forma paralela, hubo en la literatura de la Bibliotecología y la Ciencia de la Información en la década de los 40 del siglo pasado, distintas aproximaciones a la utilización de la palabra “control” en forma aislada, pero expresiva. Como es sabido, el concepto de “control” se asocia con cuestiones de planificación, gestión y administración, es decir, con procedimientos racionalmente organizados que pretenden ser efectivos en el manejo de un segmento de la realidad. Controlar implica incidir desde una posición de gobierno o dominio, sobre la naturaleza o sobre determinados procesos biológicos o de la vida en sociedad, para alcanzar una correcta fiscalización, el desarrollo de políticas dirigidas a

la obtención de resultados previstos, o una gestión orientada a la excelencia. De allí que en diferentes ámbitos se hable de “control de natalidad”, “control en la venta de psicofármacos”, “controles sanitarios” o “control de calidad.”

Una sencilla búsqueda en la colección disponible en línea de *College & Research Libraries* (disponible en <http://crl.acrl.org/>), una de las más antiguas publicaciones periódicas de Bibliotecología todavía existentes, demuestra que ya desde la creación de la revista en 1939, la palabra inglesa *control* era utilizada en forma regular en distintos textos del área, algunas veces con su sentido general, y otras en expresiones que avizoraban su incorporación al vocabulario técnico de la especialidad, como en *administrative control*, *cooperative control*, *methods of control*, *library controls*, *bibliographical control*, expresiones presentes en diversos artículos entre 1939 y 1942.

El abstract de un documento del año 1961, tomado de la revista *American Documentation*, ofrece en su breve contenido una constelación de términos derivados o relacionados que contienen, todos ellos, la palabra control: *vocabulary treatment and controls*, *syntactical control techniques*, *controls for providing automatic indexing or search* (COSTELLO JR., 1961). Estos ejemplos tomados al azar, dan cuenta del fuerte vínculo que se había establecido ya entre la herramienta racionalizadora y normalizadora del control y los procesos de almacenamiento y recuperación de información.

En esta secuencia de acontecimientos, no parece desatinado sostener que la palabra *control*, ampliamente utilizada como elemento común en la terminología de la época y frecuentada por contagio por autores de diversos países, haya sido la llave para el traslado del término *vocabulary control* desde el campo más o menos contiguo de la Enseñanza de Lenguas hacia la Bibliotecología. Es así que en el mismo amanecer de la década de 1950 comienza a aparecer el término *vocabulary control* en distintos documentos. La exhaustiva revisión de literatura realizada ofrece como mejor

prueba de su primera cita hasta el momento, un informe de un Comité sobre Organización de la Información, que se publicó en el primer número de la revista *American Documentation* (BALL, 1950). El mismo Comité, pocos meses después, utiliza en una reseña bibliográfica la expresión *controlled vocabulary*, lo que demuestra que por esa época ambos términos (vocabulario controlado y control de vocabulario), adquirirían formalización y carta de ciudadanía en Bibliotecología (COMMITTEE ON ORGANIZATION OF INFORMATION, 1950).

Al generalizarse y aceptarse paulatinamente, el término *vocabulary control* fue sustituyendo a expresiones como *normalization* o *standardization*, que habían sido utilizadas para caracterizar las prácticas relacionadas con la selección formal de encabezamientos de materia desde principios del siglo veinte (BLACK, 1946), tales como la determinación de términos preferidos, la indicación de notas de alcance y la remisión entre sinónimos. De hecho, algunas listas de encabezamientos eran también nominadas por entonces, como *standard lists* (SHIELDS, 1945).

Curiosamente, la visibilidad de la expresión *vocabulary control* o sus equivalentes en otros idiomas nunca ha sido especialmente significativa en los títulos de libros y publicaciones del área de Organización del Conocimiento. Resulta ilustrativo comprobar que en las más de las 170 referencias consignadas en la bibliografía correspondiente a la segunda edición de la obra *Vocabulary control for Information Retrieval* (LANCASTER, 1986), apenas una incluye explícitamente en su título la expresión *vocabulary control*. Shapiro (1989), autor de un bien fundamentado repertorio de terminología bibliotecaria, ni siquiera menciona las expresiones *vocabulary control* o *controlled vocabulary* en su obra.

En lo que hace a la Organización del Conocimiento, la palabra “control” continuó siendo utilizada en los últimos cincuenta años en dos direcciones

principales: (a) por un lado, vinculada a los procesos de normalización de la forma gráfica de los términos de indización y de las relaciones establecidas entre ellos, genéricamente denominados como procesos u operaciones de “control de vocabulario”; (b) por otro lado, asociada a la identificación de los repertorios surgidos de esos procesos, que comenzaron a denominarse “listas controladas”, “índices controlados”, “lenguajes controlados” y finalmente “vocabularios controlados”, expresión esta última, que parece haberse establecido de modo definitivo en la terminología del área, sin perjuicio de la aparición esporádica de las otras expresiones alternativas.

No debe olvidarse que la historia del término *vocabulary control* se asocia indisolublemente con el surgimiento y desarrollo de los primeros tesauros (AMERICAN INSTITUTE OF CHEMICAL ENGINEERS, 1961; U.S. DEFENSE DOCUMENTATION CENTER; 1962; U.S. NATIONAL INSTITUTES OF HEALTH, 1963, entre otros). Los tesauros pioneros intentaron superar las dificultades planteadas por el Sistema Uniterm introducido en 1951 por Taube (1955). Este sistema representaba los asuntos de materia de los documentos mediante palabras simples, que Taube bautizó como unitérminos, y que se registraban en fichas. Los unitérminos eran expresiones del lenguaje natural que rápidamente exhibieron sus debilidades: dificultad para expresar conceptos que requerían combinaciones de palabras, homonimia, alto número de sinónimos, etc. Taube influyó en el surgimiento de los tesauros de dos maneras: la primera, a contracorriente, provocando la necesidad de formas controladas del vocabulario; la segunda, “en el gran número de términos constituidos por palabras simples que pueden encontrarse en los primeros tesauros y en los primeros estándares para la construcción de tesauros” (LANCASTER, 1986, p. 31). El historial del proceso de creación de los tesauros y su inserción en la cronología de los estudios de información, puede consultarse en detalle en dos fuentes acreditadas (ROBERTS, 1984; ROE; THOMAS, 2004).

FUNCIONES DEL CONTROL DE VOCABULARIO

Los estándares constituyen las fuentes más autorizadas para el análisis de diferentes cuestiones vinculadas con el control de vocabulario, pues son los responsables del establecimiento de principios y criterios generales, así como de normas específicas en la materia. Proponen además reglas de aplicación universal y buscan unificar criterios metodológicos para desarrollar tesauros y otros vocabularios controlados, y favorecer la interoperabilidad entre vocabularios.

Un estado actual de la vigencia, cobertura y relevancia de los diferentes estándares de circulación internacional puede verse en el trabajo de Sánchez Cuadrado; Colmenero Ruiz y Moreira (2012).

Del análisis comparado del contenido de tres de los estándares de mayor prestigio en el mundo (BRITISH STANDARDS INSTITUTION, 2005; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010; INTERNATIONAL STANDARD ORGANIZATION, 2011), así como de los apuntes de autores que han incursionado en este tema, puede afirmarse que el control de vocabulario cumple con cinco funciones diferentes y complementarias:

1. *Restricción.* La restricción es una de las funcionalidades más destacadas en los procesos de control de vocabulario, bajo la idea de que cuanto más acotado sea el universo conceptual que se delimita, más posibilidades existen de generar formas consistentes de indización, búsqueda y acierto, y de guiar intuitivamente al navegante dentro de un mar de documentos y recursos. Habitualmente se establecen límites en cuatro direcciones, a saber:

i) la cantidad de términos a ser incluidos. Como afirma LANCASTER, el vocabulario debe integrarse con un “conjunto limitado de términos” (LANCASTER, 1986, p. 3). Se pretende dar la mayor cobertura posible de un área del conocimiento con la menor cantidad de expresiones significativas. Algunos factores inciden para que esta restricción

se atenúe o se excluya: la vocación de recopilación exhaustiva de la terminología de un campo temático, la necesidad de indizaciones específicas y muy específicas, y la dimensión del área que se está controlando, entre otros elementos a considerar.

ii) la extensión de cada término. Un término de indización no debería sobrepasar las tres palabras, salvo excepciones, pues su extensión podría afectar en alguna medida el acceso y recuperación de un documento. “Por regla general, cuanto mayor es el número de palabras que compone un descriptor, mayores son las posibilidades de que en entornos electrónicos, se pueda recuperar información con *ruido*” (entendido como conjunto de documentos, referencias o recursos de información que resultan ser no pertinentes o no relevantes a la búsqueda, al momento de la recuperación), “o de *fuga de datos* (pérdida o invisibilidad de información pertinente a una demanda y que no es recuperada a partir de las claves temáticas del usuario demandante)” (SIMÓN et al., 2013, p. 5). Cabe agregar que cuanto más extensa es la formulación de un término, más posibilidades existen de que se cometan errores ortográficos o surjan variantes lexicales del mismo, por derivación o deformación de la expresión inicial.

iii) el alcance conceptual de cada término. Como establece un criterio tradicional en la materia “el alcance de los términos se restringe deliberadamente de acuerdo con significados seleccionados” (INTERNATIONAL STANDARD ORGANIZATION, 2011, p. 16). Se tiende a asignar a cada término un solo significado, excluyendo las demás acepciones, en cumplimiento del principio de univocidad (un término – un significado, y viceversa). Por ese motivo, el estándar mencionado agrega que “a diferencia de los términos de un diccionario, que pueden estar acompañados de un número de definiciones diferentes que reflejan el uso común, cada término de un tesauro generalmente está restringido al único significado que satisface más efectivamente las necesidades de un sistema de información.” (INTERNATIONAL STANDARD ORGANIZATION, 2011, p. 16).

iv) la cantidad de relaciones que se establece entre un término y otros afines. En los vocabularios controlados se limitan las relaciones a un número suficiente para establecer al menos los vínculos paradigmáticos entre diferentes conceptos. Por otra parte, dentro del número de relaciones, se aspira a un equilibrio entre sus distintos tipos (jerárquicas, enumerativas, asociativas), a través del cual se brinde una idea aproximada de la situación del término en la estructura conceptual de referencia. Desde la perspectiva restriccionista, el control de vocabulario opera como una expresión de economía lingüística.

2. *Desambiguación.* La homonimia y la polisemia, las variantes, las sinonimias y/o cuasi sinonimias son manifestaciones habituales de la ambigüedad del lenguaje, provocando dudas, solapamientos y confusiones en torno al significado de los términos propios de una disciplina. Las ambigüedades provocadas por las formas sinonímicas y las variantes se esclarecen a través de relaciones de equivalencia. Según Bakhtin, “en el plano ideológico, la palabra es una unidad neutra, apta para adecuarse a diferentes padrones culturales. Y eso ocurre, porque ella es portadora de una gama de significación que la torna capaz de asumir sentidos o valores diferentes, dependiendo del contexto” (BAKHTIN apud CINTRA et al., 2002, p. 67-68). Si como afirma Moreiro (1994, p. 207), “los términos controlados se aplican para caracterizar contenidos en los documentos”, las técnicas de desambiguación procuran la mayor precisión respecto al alcance conceptual y al sistema de relaciones de cada término. Los métodos habituales de desambiguación son los siguientes:

i) Aplicación del principio de univocidad. En un vocabulario controlado los términos polisémicos solo conservan uno de sus significados, excluyéndose todos los demás.

ii) Aclaración del campo temático o del contexto de uso de un descriptor que tiene homónimos. Este procedimiento se realiza aplicando calificadores, que son entendidos como “expresiones que se usan para distinguir entre los diferentes significados o

niveles de aplicación de los homónimos, orientando de esa manera la indización. Habitualmente el calificador aparece integrado al término principal en alguna modalidad fácilmente perceptible; por ejemplo, entre paréntesis como en *Dilatación (Termodinámica)*.” (BARITÉ et al., 2013, en línea)

iii) Indicación de notas de alcance, notas de definición y notas históricas. Estas notas permiten restringir, extender, aclarar, definir o explicar el uso de un término en un vocabulario controlado. Por ese motivo, buscan diluir toda ambigüedad de sentido, pertinencia o alcance. Estas notas se tratarán más adelante.

iv) Indicación del sistema de relaciones de cada término. Una modalidad indirecta de desambiguación se produce a través de la interpretación del significado de un término a partir de su situación relativa en la estructura conceptual que sostiene al vocabulario controlado. La identificación del término tope, el término genérico, el o los términos específicos y relacionados, y cualquier otra forma de conexión de un término con otros permiten inducir algunos atributos o características del objeto cuyo concepto es representado. Así, si el descriptor ÁGUILAS tiene como término genérico al descriptor AVES, y como término tope al descriptor VERTEBRADOS, se puede establecer que las águilas son aves vertebradas.

Si bien los autores coinciden en este procedimiento general para el control de vocabulario (CAVALCANTI, 1978; GIL LEIVA, 1999; CINTRA et al., 2002; MARTÍNEZ; VALDEZ, 2008), pueden requerirse estrategias distintas según la disciplina de que se trate, ya que no todas ellas tienen la misma consistencia terminológica, en lo que hace a la estabilidad de las relaciones internas de sus conceptos (MOTE, 1962; BHATTACHARAYYA, 1974; SVENONIUS, 1986).

3. *Normalización.* Por sus características y objetivos, el control de vocabulario se aboca al establecimiento de formas estandarizadas de representación del conocimiento. En esta medida se vincula con

procedimientos de control de autoridades. Es posible considerar, inclusive, que el control de vocabulario es la vertiente del control de autoridades que corresponde a la normalización de autoridades temáticas. La normalización comprende la fijación de forma y de sentido, y la situación ponderada de cada término en la estructura conceptual de pertenencia, tendiendo a una mayor coincidencia entre la indización y las búsquedas de los usuarios (AITCHISON; GILCHRIST; BAWDEN, 2000). Los estándares acostumbra sugerir un conjunto importante de reglas formales de normalización, para tomar decisiones en torno, por ejemplo, a: elección de singulares y plurales; ortografías correctas o sugeridas; uso de mayúsculas, siglas, abreviaturas, acrónimos, adjetivos, adverbios, iniciales y signos gráficos; utilización de nombres populares o científicos; elección de nombres propios de personas, lugares e instituciones.

La consecuencia del establecimiento de reglas precisas para la normalización determina que todos y cada uno de los términos que integran un vocabulario controlado, han pasado por procesos de depuración consistentes y preestablecidos.

4. *Jerarquización.* Si bien los sinónimos y cuasi sinónimos son intercambiables en un discurso o en la comunicación habitual, en la mayoría de los sistemas de organización de conocimiento solo uno de ellos es reconocido como término autorizado o preferido, mientras los demás conservan solamente valor de referencia, y aún otros son dejados de lado. Así, la distinción entre términos autorizados, no autorizados y excluidos, es constante en el desarrollo de un sistema de organización del conocimiento, e implica una jerarquización de los descriptores o encabezamientos.

Jerarquizar implica valorar con distinta ponderación a los términos, y en este punto corresponde incorporar, por asociación natural, la noción de garantía. Esta es entendida como el “principio de autoridad que legitima la inclusión o exclusión de términos dentro de un sistema de organización del conocimiento, así como de las relaciones que se determinan entre dichos términos” (BARITÉ

et al., 2013). Son varias las modalidades de justificación que se han agregado a la garantía literaria propuesta en forma pionera por Hulme hace más de un siglo (HULME, 1911): garantía de usuario (LANCASTER, 1977), académica (SACHS; SMIRAGLIA, 2004), organizacional (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010), cultural (LEE, 1976; BEGHTOL, 2002), entre otras.

Si se utilizara la garantía literaria como respaldo, debería considerarse término preferido aquel que fuera más ampliamente consignado en la literatura. Por el contrario, si se tomara como base la garantía de usuario, deberían tomarse en cuenta las expresiones mayormente utilizadas por los consultantes de un sistema de información en sus demandas presenciales o remotas. Si se entendiera pertinente aplicar la garantía académica para establecer los términos preferidos habría que apelar a la opinión experta de los especialistas en el área de que se trate. Y si se tratara de una herramienta de uso corporativo, como un tesoro para una empresa industrial con productos y especificaciones propios, podría considerarse como justificación la garantía organizacional, que comprende al conjunto de expresiones efectivamente utilizadas hacia adentro de la empresa y hacia sus clientes, en los procesos de producción, comercialización y/o comunicación de bienes y servicios.

El apoyo en un tipo de garantía contribuye a establecer criterios uniformes y racionales para la selección y jerarquización de terminología y permite realizar por lo menos cuatro afirmaciones respecto a los términos de indización elegidos: a) el término existe; b) corresponde a un concepto que es reconocido por los especialistas del área y/o por los usuarios; c) es el más representativo respecto a otros que corresponden al mismo concepto; d) cuenta con una garantía que lo respalda. Desde esta perspectiva, la jerarquización es una modalidad de validación de la terminología.

5. *Relación.* El control de vocabulario trabaja sobre dos vínculos: las relaciones intra-término, donde se dilucida la situación de sinónimos, cuasi sinónimos

y variantes; y las relaciones entre términos, que tradicionalmente se establecen bajo distintas modalidades de relación jerárquica o asociativa (CINTRA et al., 2002; AITCHISON; CLARKE, 2004; NAUMIS, 2007). Estructuras nuevas como las ontologías permiten diversificar y enriquecer el sistema de relaciones, pues se apoyan en vínculos funcionales o pragmáticos que se corresponden con los procesos de la vida real.

Las relaciones entre términos reguladas por los estándares y establecidas en los manuales corrientes (AITCHISON; GILCHRIST; BAWDEN, 2000; CURRÁS, 2005) se han mantenido estables a lo largo del tiempo, y son: de equivalencia o de identidad, jerárquicas y asociativas. A ellas, la norma ISO 25.964-1 agrega las relaciones personalizadas (*customized relationships*), que dicho estándar ejemplifica con la relación entre una sigla y el nombre desarrollado de una institución (INTERNATIONAL STANDARD ORGANIZATION, 2011, p. 67). No obstante, las relaciones personalizadas pueden ser consideradas como una variedad de las relaciones de equivalencia o identidad. Más adelante se hará referencia a los tipos clásicos de relación entre términos.

Los estándares colocan diferentes énfasis respecto a las funciones del control de vocabulario. Así, el estándar norteamericano pone el acento en el proceso de organización de una lista de términos, a partir de operaciones de desambiguación, mencionando también la jerarquización (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010, p. 10). El estándar británico prioriza la restricción, pues ella puede “aumentar la posibilidad de que los indizadores y buscadores encuentren el mismo término representando al mismo concepto” (BRITISH STANDARD INSTITUTION, 2005, p. 6). Por último, la norma ISO 25.964-1 establece un equilibrio entre la desambiguación y la restricción, y menciona lateralmente la tarea de jerarquización entre términos (INTERNATIONAL STANDARD ORGANIZATION, 2011, p. 12).

PROCEDIMIENTOS DE CONTROL DE VOCABULARIO

Las operaciones y los procedimientos típicos de control de vocabulario han sido recurrentemente tratados en la literatura del área, aunque con escasos apuntes críticos. En este trabajo se intenta a continuación una sistematización un poco más afinada de los mismos, y se agregan comentarios cuando así se entiende pertinente.

Son formas de control de vocabulario:

I) Control de sinónimos y variantes. Esta forma de normalización “intensifica el proceso de control sobre la variación de significado, permitiendo mayor rigor en el tratamiento de la información y [mayor] eficacia al momento de la recuperación” (CINTRA et al., 2002, p. 75). Más allá de las disquisiciones que realizan los lingüistas acerca de la existencia misma de los sinónimos (SALVADOR, 1985; DUCROT; TODOROV, 1998), lo cierto es que en los hechos existen términos cuyos sentidos pueden ser similares o, al menos intercambiables para los usuarios de un sistema de información. Por otra parte, razones de economía de recursos lingüísticos favorecen también este tipo de control de vocabulario. En este proceso se normalizan tanto formas sinonímicas más o menos aceptadas, como cuasi sinónimos y variantes de distinta índole.

Las variantes son “cada una de las formas que puede asumir un término, como resultado de los diferentes usos que una sociedad hace del mismo en razón de su diversidad lingüística, socioeconómica, geográfica, etnográfica o cultural, entre otras” (BARITÉ et al., 2013). Así es posible identificar variantes gráficas, lexicales, morfosintácticas socioprofesionales y topolectales entre otras (FAULSTICH, 1995).

La primera fase de este control consiste en la detección y el despliegue de las diversas formas gráficas bajo las cuales se representa un concepto (GARSHOL, 2004). Una de las herramientas novedosas para representar el conjunto de formas sinonímicas de

un concepto es el anillo de sinónimos, el cual es definido como “un grupo de términos que son considerados equivalentes para los propósitos de la recuperación” (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005, p. 9). Por sus características, los anillos no establecen jerarquizaciones entre los sinónimos. Cumplen dos funciones básicas: servir como términos de búsqueda a texto libre en lenguaje natural y recolectar el conjunto de sinónimos que pueden ser importantes para su representación en un sistema de organización del conocimiento.

El control de sinónimos, cuasi sinónimos y variantes involucra las siguientes etapas:

- La identificación de las formas más o menos equivalentes en la literatura más aceptada de la especialidad de que se trate.
- La selección y fijación de una de las formas como término preferido a partir de criterios preestablecidos, el cual se transformará en el término autorizado para la indización.
- La indicación de las formas no preferidas, que serán consideradas como términos no autorizados, y que constituirán igualmente vías indirectas de acceso a una colección de documentos o de recursos de información.
- La elaboración de un sistema de reenvíos desde los términos no autorizados al término autorizado. El reenvío se efectúa mediante referencias de véase en las listas de encabezamientos, a través del índice en los sistemas de clasificación, y de las abreviaturas USE/UP (use/use para) o USE/UF (*use/use for*) en los tesauros.

El control de sinónimos, cuasi sinónimos y variantes se procesa a través del establecimiento de relaciones de identidad o equivalencia. Algunas relaciones de jerarquía son tratadas como equivalentes por motivos prácticos, y referencias cruzadas que se realizan desde elementos individuales que integran un término compuesto hacia el término compuesto correspondiente (como en el caso “Arcos” USE “Arcos y flechas”).

II) Control de equivalencias. Las equivalencias son los términos que representan aproximadamente al mismo concepto en otra lengua. Entendidos los conceptos como expresiones más o menos universales, es natural que sus representaciones varíen de idioma en idioma. Así, la noción de ‘unión aduanera’ no cambia sustancialmente aunque su representación lingüística se exprese en inglés, en portugués o en español. Las equivalencias de denominaciones y conceptualizaciones entre una lengua y otra preocupan especialmente a los traductores, quienes deben hacerse cargo de las discordancias conceptuales entre comunidades lingüísticas con fondos culturales diferentes. No obstante, la necesidad de establecer y controlar equivalencias alcanza también a la Bibliotecología y la Ciencia de la Información en distintas áreas, como el trabajo de referencia, la preparación de traducciones de tesauros y la indización de documentos en lengua extranjera.

El control de equivalencias se expresa a menudo en la preparación de índices de equivalencias, que se integran como anexos de tesauros, taxonomías o listas, y son especialmente útiles por su alto valor de referencia. La formalización del control es relativamente sencilla y da por sobreentendido el mismo alcance conceptual para los términos en distintos idiomas. Las listas LEMB (INSTITUTO COLOMBIANO PARA EL FOMENTO DE LA EDUCACIÓN SUPERIOR, 1985), por ejemplo, establecen la relación entre equivalencias en su anexo, como sigue:

Abduction

Rapto

Aberration

Aberración (Astronomía)

La siguiente presentación controla equivalencias al tiempo que determina la calidad de cada término en español:

Literary warrant**Garantía literaria** (descriptor)

Respaldo en la literatura (no descriptor)

Justificación en la literatura (no descriptor)

La cualidad de descriptor o término autorizado puede ser marcada tipográficamente por un tipo de letra diferente (negrita en este caso), o por especificaciones entre paréntesis o de alguna otra manera expresiva.

El control de equivalencias constituye una modalidad de control de vocabulario porque además de establecer las nominaciones más convenientes para el mismo concepto en dos o más idiomas, requiere un trabajo de normalización formal y semántica en todas las lenguas involucradas.

III) Control de homonimia y polisemia. La homonimia es vista como la similitud formal de palabras que son diferentes, y que remiten “a dos realidades diversas, sea por identidad fónica (homofonía) o identidad gráfica (homografía)” (CINTRA et al., 2002, p. 71). En el caso particular de los homógrafos, pueden plantearse situaciones de ‘ruido’ al momento de la recuperación de la información, por lo que las operaciones de control de vocabulario tienden a desambiguar la situación de cada homógrafo respecto a su propio contexto.

La frontera entre homógrafos y términos polisémicos no siempre es clara. Para Maniez, “los términos homógrafos son distintos por su origen”, y esa diferencia se puede rastrear en la etimología de cada uno de ellos (MANIEZ, 1993, p. 199). De este modo serían términos homógrafos ‘tradición’ como el conjunto de valores culturales compartidos y ‘tradición’ en el sentido jurídico de la entrega efectiva de un bien. Quiere decir que los homógrafos no comparten un mismo fondo semántico, y por tanto, la coincidencia en la forma de las palabras se da de manera fortuita.

La polisemia aporta otra perspectiva, pues se dice que existe cuando un término comporta más de un significado, o cuando se dan extensiones de un significado original.

Se trata de un fenómeno que se da muy frecuentemente, a partir de la aplicación recurrente de las analogías y metáforas para describir o explicar objetos y fenómenos nuevos a partir del conocimiento anteriormente aceptado (WAY, 1991; STAMBUK, 1998).

Al igual que en el caso de los homógrafos, los términos polisémicos suelen engendrar ‘ruido’ documental (GIL URDICIAIN, 1996). Este ruido es neutralizado o reducido en ellos mediante diversos procedimientos: en los sistemas de clasificación, la polisemia se expresa en los índices, y es controlada mediante la identificación de la disciplina entre paréntesis. En los tesauros, la ubicación relativa del término polisémico en sus diversas áreas permite discriminar cada sentido en función del contexto disciplinario. Así, el término ‘Virus’ podrá aparecer bajo Medicina y bajo Informática, y esas ubicaciones le darán a la expresión el sentido y el lugar correspondientes.

Los calificadores constituyen una solución clásica para aclarar el sentido de términos homógrafos o polisémicos (CAVALCANTI, 1978; BUCHAN, 1989). En estos casos, el calificador puede aparecer integrado al término principal o marcado tipográficamente o de otra manera; por ejemplo, entre paréntesis:

ESTRÉS (FISIOLOGÍA)

ESTRÉS (PSICOLOGÍA)

IV) Control de abreviaciones. Los tipos más comunes de abreviaciones son las abreviaturas, las siglas y los acrónimos. Este tipo de expresiones ofrecen diferentes dificultades, que podrían afectar la recuperación de información. Entre ellas se encuentran las siguientes:

- la misma abreviación puede corresponderse con diferentes conceptos, entidades u organizaciones;
- una expresión puede abreviarse de diferentes formas;
- las abreviaciones pueden variar de idioma en idioma;

- a diferencia de las siglas y los acrónimos, las abreviaturas no suelen utilizarse como términos de indización;
- en el caso de organizaciones, se plantea siempre la duda razonable de si el término de indización debe ser la sigla o acrónimo, o la denominación extendida de la organización.
- muchas organizaciones cambian de denominación a lo largo del tiempo.

V) *Notas de alcance*. Las notas de alcance (llamadas ‘*scope notes*’ en inglés) constituyen la solución tradicional en materia de tesauros y otros vocabularios controlados, para resolver problemas de desambiguación semántica o para situar el sentido particular de un término en un contexto. Se identifican mediante las abreviaturas SN en inglés y NA en español.

Las notas de alcance se utilizan “para restringir o expandir la aplicación de un término, para distinguir entre términos que tienen significados solapados en el lenguaje natural o para proporcionar una sugerencia al indizador o al usuario que busca información” (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010, p. 22). También se elaboran cuando se requiere especificar el sentido de un término en la estructura conceptual. Las notas de alcance solo se utilizan en casos en los que puedan surgir dudas razonables respecto al significado de una expresión, pues en general se entiende que el usuario del vocabulario controlado puede situar a un término en su contexto en función del entramado de relaciones de identidad, jerárquicas y asociativas.

Las notas de alcance se realizan para fijar un sentido exclusivo a ese término dentro del contexto temático en el cual se inscribe el lenguaje. Las razones por las que se establecen notas de alcance son, de hecho, muy variadas y han sido sistematizadas por Aitchinson, Gilchrist y Bawden (2000) y otros autores de esta manera:

- como aclaración de significado (BUCHAN, 1989);
- como indicación del rango de tópicos cubiertos. Cuando en el vocabulario se incluye sólo el

término genérico, la nota de alcance identifica los términos específicos como referencia (LEATHERDALE, 1980);

- como señal de la existencia de indicadores clasificatorios (también llamados términos falsos o postizos), que son inadecuados para la indización pero se incluyen en la sección sistemática de algunos tipos de tesauros para indicar el criterio lógico de clasificación; por ejemplo, ‘personas según su edad’, ‘personas según su sexo’, o ‘personas según su profesión’;
- como nota histórica del término, indicando por ejemplo, la fecha de adopción de un término, y de diversos cambios a lo largo del tiempo.
- como indicación de restricción de significado (BUCHAN, 1989; GIL LEIVA, 2008);

VI) *Notas de definición*. Una figura cercana a la nota de alcance es la nota de definición, en la cual sí se define a un concepto, y se desarrolla la acepción bajo la cual ese término ha sido aceptado y entendido en el lenguaje.

Constituye la modalidad más elaborada de desambiguación, si bien es la menos explorada hasta el presente, en la tradición de los vocabularios controlados. En este sentido, surgen regularmente en la literatura opiniones favorables a la inclusión sistemática de definiciones en los sistemas de organización del conocimiento (BUCHAN, 1989; HUDON, 1996; JORNA; DAVIES, 2001), especialmente en las áreas sociales y humanas, debido a la imprecisión o reformulación de su terminología (AITCHINSON, GILCHRIST; BAWDEN, 2000).

En la tradición de los sistemas de organización del conocimiento, la inclusión de definiciones para los términos o las notaciones ha sido antes la excepción que la regla. De entre los sistemas de clasificación de uso más extendido, destacan las últimas versiones de la Clasificación Decimal Dewey, que han comenzado a quebrar esa tradición, incorporando esporádicas y breves definiciones, aunque a menudo no pasan de ser notas indicativas. Las listas de encabezamientos incorporan definiciones solo cuando los términos son extremadamente ambiguos o pueden tener distinto significado en distintas áreas.

Una práctica alternativa a la elaboración de definiciones originales es el enlace de un tesoro en línea con diccionarios que se encuentran también disponibles (McILWAINE, 1998), aunque la limitación de esa metodología está en que, por una parte, en esos diccionarios los términos no están desambiguados y, por el contrario, exhiben todo su despliegue polisémico; y por la otra, en que no siempre son compatibles las definiciones de diccionarios con el alcance particular que se le quiere dar a un término en un tesoro.

VII) *Redacción de notas históricas.* Estas notas pueden registrar la fecha de introducción del término al área o al vocabulario controlado, o pueden orientar respecto a la búsqueda por términos alternativos que fueron utilizados antes en el tiempo para representar el mismo concepto. Como señala la norma ISO 25.964-1, las notas históricas se vuelven necesarias “cuando un tesoro es actualizado, y los cambios en los términos pueden afectar la recuperabilidad de un concepto” (INTERNATIONAL STANDARD ORGANIZATION, 2011, p. 23).

VIII) *Control de relaciones jerárquicas.* Una relación jerárquica se establece entre un concepto general o supraordenado y otro concepto que se subordina al primero y queda incluido en el mismo, como en los casos de ROCAS-BASALTO o BIBLIOTECAS-BIBLIOTECAS PÚBLICAS. La relación de jerarquía tiene una larga historia en la evolución del conocimiento y el pensamiento científico. Se debe a Porfirio, un filósofo griego del siglo tercero, el desarrollo de la primera clasificación jerárquica, antecedente de las taxonomías que tendrían en los siglos siguientes tan principal importancia en la sistematización de las ciencias (KWASNIK, 1999). En su obra *Isogage*, expuso una estructura jerárquica, conocida en adelante como *Arbor porphyriana* o Árbol de Porfirio, consistente en la desagregación de conceptos, partiendo de los más generales a los más simples, que es ampliamente conocida como antecedente de toda clasificación jerárquica.

Como señala Kwasnik, “las jerarquías son útiles para entidades que están bien definidas y tienen fronteras de clase claras”, aunque requieren para su

formulación “alguna teoría o modelo (...) para guiar la identificación de entidades, las reglas de asociación y distinción, y el orden en el cual esas reglas son invocadas” (KWASNIK, 1999, p. 30).

Una de las cualidades por las cuales los sistemas decimales, a partir de Dewey, tuvieron tanta aceptación es su condición jerárquica, asociada naturalmente a la decimalidad. En el sistema de Dewey, la jerarquía se expresa a través de la notación (cada número está contenido en el anterior como en la serie 444 – 444.1 – 444.12 y así sucesivamente) y de la estructura (cada una de las diez clases principales se subdivide en otras diez y así sucesivamente). Como se señala en la introducción de una de sus ediciones, “cada tópico está subordinado y a su vez es parte de todos los tópicos más amplios que lo preceden. El corolario es igualmente verdadero: lo que es válido para el todo lo es para sus partes” (SISTEMA ..., 1995, p. xxxvi-xxxv).

Las relaciones jerárquicas pueden ser de tres tipos, y las tres tienen una importante consagración en la literatura:

i) *relación genérica o de género / especie.* Es aquella que se establece entre un concepto que representa al género, y otros u otros que representan a especies o variedades del mismo. Cada una de las nociones específicas posee todas las características de la noción genérica más al menos un atributo suplementario y distintivo. Los atributos distintivos, a su vez, permiten diferenciar una especie de otra en el mismo nivel de abstracción; por ejemplo: IDIOMAS-INGLÉS; IDIOMAS-PORTUGUÉS.

ii) *Relación partitiva o de todo / parte.* En esta relación entre conceptos uno representa una totalidad o un objeto, y otro u otros representan las partes o componentes de ese objeto. Por tanto, todos los enunciados o atributos del ‘concepto parte’ están presentes en los enunciados del ‘concepto todo’, pero no al revés. Conforme a lo expuesto, la noción de parte depende de la noción del todo, de modo que no puede definirse la parte si no se define antes el todo; por ejemplo: ÁGUILAS-ALAS; PALOS-MÁSTILES.

iii) *Relación enumerativa o de casos particulares.* Es la que se establece entre un término genérico y un objeto particular o un ejemplo individual del género, que se identifica con un nombre propio. Quiere decir que el caso particular se determina perfectamente, como en los ejemplos RÍOS-AMAZONAS o MONTAÑAS-KILIMANJARO. La necesidad de una mayor precisión en la indización obliga a incluir los casos particulares cada vez en mayor número en los esquemas de los sistemas de organización del conocimiento.

Las relaciones de jerarquía de cualquier tipo se visualizan en los tesauros a través de las abreviaturas TT- TG-TE (término tope, término genérico, término específico), o sus equivalentes en inglés TT-BT-NT ('*top term*', '*broader term*', '*narrow term*')

IX) *Control de relaciones asociativas.* También llamadas de afinidad o de asociación, son aquellas que se establecen entre dos o más asuntos por algún tipo de proximidad semántica, conceptual o funcional entre sí, con exclusión de las relaciones jerárquicas o de sinonimia. Los términos involucrados en una relación asociativa tienen una conexión recíproca y son vinculados habitualmente por las personas y por la documentación. La fuerza de su conexión está dada por la estabilidad del vínculo. En los tesauros este tipo de relación se expresa en la abreviatura TR ('término relacionado') o su equivalente en inglés RT ('related term') (BRITISH STANDARDS INSTITUTION, 2005; MARTÍNEZ; VALDEZ, 2008).

Existe solo un consenso parcial en la literatura acerca de los distintos tipos de relación asociativa que se pueden establecer entre conceptos. La recurrencia en el estudio de estas relaciones puede interpretarse como una tentativa por "descubrir si existe un conjunto universal de relaciones [conceptuales] aplicables a través de todos los contextos" (McILWAINE, 1998, p. 145), que pase por encima de las diferencias sociales y culturales. Las variedades más habituales de asociación entre asuntos son las siguientes:

Relación asociativa general. Es la que se establece entre dos asuntos cualesquiera con cierta regularidad, ya que el vínculo está instalado en la realidad; por ejemplo, 'AJEDREZ-CAMPEONATOS'; 'POKER-DINERO'.

Relación de coordinación o derivacional. La que se establece entre dos términos coextensivos, es decir, que pertenecen al mismo conjunto por compartir uno o más atributos; por ejemplo: 'LIBROS IMPRESOS'; 'LIBROS ELECTRÓNICOS' (FOSKETT, 1996; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010).

Relación causa-efecto o causal. La que se establece entre dos términos, uno de los cuales corresponde a una causa y el otro, a un efecto o consecuencia. Puede incluir, además, al nexo causal que permite este vínculo; por ejemplo, 'ACCIDENTE-VÍCTIMA'; 'ENSEÑANZA-APRENDIZAJE', 'VIRUS-INFECCIÓN'. (FOSKETT, 1996).

Relación de oposición o antonimia. La que se da entre dos términos, denominados antónimos, cuyos enunciados se excluyen y oponen recíprocamente; por ejemplo, la relación 'TRABAJO REMUNERADO'- 'TRABAJO NO REMUNERADO'.

Relación genética. Es un tipo de relación causal en la cual un objeto precede a otro, en el tiempo o de alguna otra forma, y provoca el surgimiento del otro; por ejemplo, las relaciones 'ÁRBOL-HOJAS', 'PADRES-HIJOS', 'SEMILLA-PLANTA' (FOSKETT, 1996).

Relación entre objetos y sus propiedades. Por ejemplo: 'PIEL'- 'PIGMENTACIÓN'.

Relación intra-disciplinaria. Es aquella que se establece entre una disciplina y su objeto de estudio; por ejemplo: 'SOCIOLOGÍA-SOCIEDAD' (CURRÁS, 2005).

Relación ocupación-persona. Es la que se manifiesta entre una determinada ocupación y las personas que la practican; por ejemplo: 'EDUCACIÓN-PROFESORES' (CURRÁS, 2005).

Relación instrumental o proceso-agente. Es la relación que se caracteriza porque uno de los objetos necesita del otro para realizarse o cumplir su función específica; por ejemplo, las relaciones ‘PINTURA-PINCEL’ y ‘LUBRICANTE-ENGRASE’ (CURRÁS, 2005).

Las nueve modalidades que se han reseñado son las que se consideran tradicionalmente como prácticas sistemáticas de control de vocabulario, para la construcción o revisión de un sistema de organización del conocimiento. La mayor o menor rigurosidad con que se consideren o apliquen estas modalidades repercute directamente en la calidad del sistema de que se trate. El estándar ANSI/NISO Z39.19 para la construcción de vocabularios controlados monolingües, por ejemplo, determina que las listas son las estructuras menos complejas porque se apoyan en un control de vocabulario limitado, y son seguidas en un eje de complejidad creciente por los anillos de sinónimos, las taxonomías y los tesauros (National Information Standards Organization, 2010).

CONTROL DE VOCABULARIO Y ENTORNO DIGITAL

Cabe preguntarse hasta dónde el control de vocabulario -en un mundo orientado a la búsqueda libre mediante Internet y dominado por las aplicaciones y herramientas tecnológicas de última generación- sigue constituyendo una herramienta necesaria para asegurar el mejor desempeño de usuarios, computadoras y bibliotecarios en la recuperación de documentos o recursos de información, o de datos puntuales contenidos en los mismos.

También corresponde interrogarse si no habría que adaptar las estructuras conceptuales, creadas con las finalidades más diversas, a las características, las competencias y la versatilidad que exhiben los nacidos digitales en sus modalidades de comunicación y transacción con la información. La referencia a los nacidos digitales implica asumir enteramente que quienes pertenecen a generaciones anteriores tendrán que adaptarse a los modelos de comunicación e información que los más jóvenes impongan. Y, en todo caso, habría que considerar qué papel juega y jugará el lenguaje natural en estos

procesos, y cuáles serán los límites impuestos por la realidad para el desarrollo de nuevos vocabularios controlados (ABD AZIZ, 2011).

En las décadas de los 80 y los 90, el avance de la computación y la paulatina construcción del universo conceptual y tecnológico de Internet provocó dos procesos aparentemente contradictorios: por una parte la web apeló a herramientas tradicionales de la Bibliotecología para organizar sus sitios web. De esta forma surgieron los directorios de buscadores, cuyas estructuras taxonómicas fueron replicadas para la organización de contenidos de sitios web institucionales de oficinas del Estado y de corporaciones, y que aún hoy se mantienen como formas de acceso alternativas a la búsqueda por lenguaje natural. En sentido contrario, se dio un claro resurgimiento del lenguaje natural como principal herramienta de búsqueda de información. Ambas tendencias fueron verificándose en paralelo (CHU, 2003). En este período importantes autores del área se mostraron francamente escépticos respecto a la utilidad de los vocabularios controlados (CLEVERDON, 1984; SALTON, 1996), al tiempo que se verificó la profundización de los estudios de *Natural Language Processing* o NLP (JACKSON; MOULINIER, 2007).

En la primera década de este siglo, se abrió un espacio para el desarrollo de modelos colaborativos, centrados en los intereses y la acción de los usuarios, a través de interfaces evolucionadas, apoyadas en el lenguaje natural para la representación y recuperación de información. El control de vocabulario siguió utilizándose en forma predominante en sistemas de información tradicionales, aunque mantuvo una perceptible influencia en las formas de tratamiento de la información llevadas adelante por los responsables de catálogos, bases y bancos de datos en línea. También influyó en los modelos de presentación de información en portales de Internet.

Es sabido por quienes han profundizado en los estudios de representación del conocimiento, que los vocabularios controlados se han nutrido siempre del lenguaje natural, tanto en su formulación inicial

como en sus revisiones y actualizaciones. Quienes controlan vocabulario no tienen mejor opción que partir del lenguaje natural, porque ello les asegura la representatividad, actualidad y adecuación de la terminología (FIDEL, 1992). Difícilmente alguien cuestione la más reconocida fortaleza del control de vocabulario: la presentación racional y organizada de términos y conceptos ofreciendo un mapa verosímil de un área del conocimiento, y la posibilidad de ordenar documentos, objetos y recursos en función de esa organización racional. Lo que está en debate no es la fiabilidad y la utilidad real de los vocabularios para organizar, sino para recuperar información.

Es posible que nos encontremos en un punto de encrucijada, en el cual “todavía necesitamos algunos ‘métodos tradicionales’ en la organización de la información, como los metadatos y la indización humana, y debemos combinarlos con los enfoques nuevos y emergentes, en especial en el contexto de un acceso integrado a la información” (SLAVIC, 2007, p. 580-581).

Hoy se discute si el estadio de desarrollo tecnológico actual se corresponde con el nivel web 2.0 o web 3.0. En cualquier caso, los rasgos de identidad que en la actualidad exhibe la *World Wide Web* son fácilmente reconocibles: la existencia de un modelo de gestión informática basada en servicios que están disponibles en una “nube”, mejor rendimiento y compatibilidad entre los programas informáticos, formas colaborativas e interactivas de construcción, registro, acceso, uso, reuso e intercambio de datos, el acceso a internet desde fuentes alternativas a las computadoras (tabletas, *smartphones*, celulares o móviles), desde cualquier lugar y en cualquier momento, la introducción de agentes inteligentes a través de los cuales se desarrollan experiencias innovadoras -como la Web 3D o la Web Geoespacial-, y el potencial de aplicación de metadatos semánticos y ontológicos que permitan una aproximación ya semántica, ya funcional, a documentos, contenidos y datos disponibles en Internet.

El vertiginoso avance de Internet y la tecnología digital “está por derrumbar (o repaginar) todo lo que conocemos...” (Da SILVA; Da SILVA, 2011, p. 114), y ha impactado en, por lo menos, tres direcciones relacionadas con el control de vocabulario:

- i) El surgimiento de nuevos tipos de sistemas controlados o semi-controlados, cuya estructuración interna pone en cuestión el modelo tradicional de creación de vocabularios controlados.
- ii) La evolución de interfaces y ayudas semi-controladas para la búsqueda de información, que operan como formas nuevas de ‘control’ lingüístico.
- iii) El desarrollo de normas y procedimientos para favorecer la interoperabilidad.

Analícemos brevemente cada uno de estos espacios, para identificar sus puntos de ruptura o de continuidad con la tradición del control de vocabulario.

Nuevos sistemas. En los últimos años se han desarrollado sistemas de organización del conocimiento originales, que buscan responder a exigencias renovadas de los usuarios. Solo se mencionarán aquí dos de ellos: las taxonomías web y las clasificaciones sociales o folksonomías.

Las taxonomías web son estructuras que contienen exclusivamente relaciones jerárquicas entre los términos. Cada término puede contener a otros más específicos, lo que se devela cliqueando sobre los mismos. En todos los casos, los taxones “están conectados mediante algún modelo estructural (jerárquico, arbóreo, facetado) y especialmente orientado a los sistemas de navegación, organización y búsqueda de los contenidos de los sitios web” (CENTELLES, 2005).

Una investigación realizada sobre distintas taxonomías encontradas en Internet, que incluyó entre otras, las encontradas en los sitios de GlaxoWellcome S.A., *British BroadCasting*

Corporation, el Servicio Postal de Estados Unidos y la por entonces denominada PricewaterhouseCoopers (GILCHRIST; KIBBY, 2000) identificó cinco modalidades taxonómicas diferentes: i) directorios web; ii) taxonomías destinadas a favorecer la indización automática; iii) taxonomías creadas por categorización automática; iv) como filtros del tipo *'front end'* es decir, como estructuras de apoyo a la formulación de la búsqueda de información o a la navegación en Internet o Intranet (véase por ejemplo, www.wordmap.co.uk); y, v) taxonomías corporativas, como formas híbridas entre tesauros y ontologías de apoyo a la búsqueda y recuperación de la información, que se han vuelto comunes dentro de sitios institucionales. El sistema de las etiquetas que representan a los términos y las denominaciones en las taxonomías de los tipos i, ii, iv y v, se respalda en mayor o menor medida, en expresiones controladas o semi-controladas.

Las folksonomías o clasificaciones sociales, por su parte, son repositorios de documentos o recursos de información, generalmente audiovisuales, que se encuentran disponibles en Internet. Cada documento o recurso recibe una o más etiquetas (*tags* o *labels*), que son asignadas libremente por quienes dejan disponible el mismo, en forma individual en la práctica, y en forma asociativa por acumulación, para categorizar contenidos tales como fotografías, videos o textos (NORUZI, 2006). Los sitios y sistemas de información web así organizados tienden a ser muy populares (por ejemplo, *You Tube*, www.youtube.com). Como señalan entusiastamente Da Silva y Da Silva, esta modalidad de indización por lenguaje natural “¿es la sabiduría de las multitudes en acción!” (Da SILVA; Da SILVA, 2011, p. 124). Solo recientemente se ha comenzado a estudiar la calidad de la indización y la recuperación de información a partir de folksonomías (PETERSON, 2006; CATARINO; BAPTISTA, 2007; HANSSON, 2014).

Evolución de interfases y ayudas. Un buen ejemplo de cómo las aplicaciones informáticas contribuyen a procesos de control está dado en las mismas folksonomías. En efecto, todo haría suponer que

las clasificaciones sociales solamente dan forma a estrategias de recuperación de información en Internet mediante el lenguaje natural proporcionado por los contribuyentes y usuarios. Sin embargo, los sistemas de ayuda matizan esa afirmación: por ejemplo, el menú de opciones que se abre a medida que se van digitando letras en la ventana (una aplicación también corriente en *google* y otros buscadores) ofrece formas normalizadas que resumen búsquedas anteriores de usuarios, y que permiten dejar de escribir si la opción buscada aparece en la lista. A medida que crece el caudal histórico de consultas, el menú de opciones comienza a parecerse bastante a un listado alfabético de entradas temáticas, tal como se acumulaba en los viejos catálogos manuales que recogían las entradas secundarias por tema. Muchas de estas clasificaciones sociales son intervenidas por analistas que eliminan errores ortográficos o inconsistencias gramaticales, y testean la funcionalidad del sistema. De ese modo, a lo largo del tiempo estas folksonomías van incrementando las ayudas semi-controladas.

La web semántica participa de este proceso aportando interfaces que permiten no solo organizar contenidos de un modo diferente, sino también proponer nuevas formas de buscar y acceder a información, que son fácilmente asequibles a usuarios con un nivel de instrucción mediana y baja. En su nivel actual de desarrollo la web semántica favorece la coexistencia de formas lingüísticas de recuperación (escribiendo en una ventana la ecuación de búsqueda) con otras basadas en el cliqueo. De este modo democratiza aún más el acceso a servicios y aplicaciones que se sustentan en la existencia de grandes bancos de información activa, actual y y sindicada. Como señala Marcondes, el uso de las tecnologías de web semántica “son propuestas para mejorar la organización de las informaciones en sitios web, para *browse* y navegación y para apoyo a la recuperación de informaciones” (MARCONDES, 2011, p. 141).

Además de los menús que se abren a medida que se escriben letras en las ventanas de búsqueda, aparecen sugerencias del tipo “quizás quiso decir”, formas de búsqueda avanzada que mejoran la recuperación de

datos, y el uso de comillas para refinar la búsqueda mediante expresiones compuestas. Nada de estas nuevas aplicaciones hará, además, que se abandone al ya “viejo” *browsing*, que constituye otra forma usual no clasificatoria de buscar información, y que a menudo incorpora términos normalizados para encaminar a los usuarios. De este modo, “todos los mecanismos de búsqueda se valen de sofisticados y complejos algoritmos para indizar y consultar la información. Diferentes empresas del segmento de tecnología de la información compiten por la innovación, buscando ser la más rápida, la de mayor base catalogada, la que ofrece mejor consulta a las necesidades de los usuarios. Al mismo tiempo en que las empresas innovan, la búsqueda de la diferenciación crea diferentes visiones sobre como la información es tratada.” (Da SILVA; Da SILVA, 2011, p. 123).

Los usuarios se han acostumbrado a buscar información sin mediación, a apropiarse rápidamente de las herramientas innovadoras y a integrarlas a su forma particular de usar el lenguaje natural. Muchos de esos usuarios van creando sus propias bibliotecas digitales, tomando recursos disponibles en Internet que están asociados a sus intereses y preocupaciones, y generando sistemas domésticos o *ad hoc* de clasificación que, al parecer, funcionan. Es posible que los de talante más racionalizador recreen formas básicas de control de vocabulario para encontrar la información que necesitan en las bibliotecas digitales que portan en su computadora o su *pendrive*. Y es que la práctica misma, y la necesidad de ahorrar tiempo y encontrar rápidamente lo que se busca es, en última instancia, el impulso que lleva a soluciones normalizadoras.

Interoperabilidad. Interoperabilidad (o interoperatividad o interoperación) es un concepto genérico que se aplica en diversos ámbitos, y cuya idea nuclear expresa la posibilidad de compatibilizar y hacer actuar de consuno sistemas diferentes que comparten, no obstante, una base común. En Organización del Conocimiento, la interoperabilidad se especifica en tres sentidos: el intercambio de datos entre vocabularios controlados, el intercambio

de datos entre un vocabulario controlado y el lenguaje natural o con instrumentos basados en el lenguaje natural (como las listas de palabras clave), y el intercambio de datos entre un vocabulario controlado y otros elementos presentes en un sistema de información.

En una situación ideal, la interoperabilidad se da relacionando vocabularios con suficiente detalle para permitir el traslado de búsquedas o la descripción de recursos desde un vocabulario que no dio respuesta a otro que es más efectivo; o la ampliación de la consulta desde un vocabulario especializado en un área de conocimiento a otro que da cobertura a esa y a otras disciplinas o áreas relacionadas. (NATIONAL INFORMATION STANDARD ORGANIZATION, 2010).

Según el estándar británico hay dos razones que demandan un proceso de interoperabilidad. Por una parte, “la tecnología permite el intercambio de datos a nivel mundial en una escala nunca antes imaginable. En segundo lugar, las presiones económicas dictaminan que los recursos de información preparados para una aplicación o contexto, pueden quedar fácilmente disponibles para otras aplicaciones”. (BRITISH STANDARDS INSTITUTION, 2005, parte 4, p. 1). En cualquier caso, la interoperabilidad requiere sistemas que compartan los mismos protocolos de computadora. La variedad y la complejidad de los modelos de interoperabilidad merecen un tratamiento particular que excede los objetivos de este trabajo. Sin perjuicio de ello, cabe preguntarse si ciertas formas de interoperabilidad operan como modalidades de control de vocabulario de segundo nivel, como un supra-control -si puede llamarse así-, que busca armonizar los procesos de control aplicados a vocabularios contruidos bajo lógicas diferentes, con vistas a favorecer el acceso a grandes continentes de información (catálogos unificados de redes de bibliotecas, redes de información, bancos nacionales e internacionales de datos, repositorios interinstitucionales, etc.).

CONCLUSIONES

Lo que conceptualmente se entiende hoy como control de vocabulario hunde sus raíces históricas en los primeros procesos de normalización de las expresiones lingüísticas que representan conceptos y sus relaciones, que se dieron ya avanzado el siglo XIX, con el desarrollo de los primeros sistemas de clasificación y las reglas pioneras para la representación temática en los catálogos. Recién a partir del segundo cuarto del siglo XX los procesos de estandarización y normalización de encabezamientos de materia y los experimentos de Taube confluyen en una práctica sistemática que da fundamento a los métodos de construcción de los primeros tesauros. En este trabajo se ha intentado dar prueba de que la efectiva entrada del término *vocabulary control* se dio desde el área de la enseñanza del inglés como segunda lengua, y que se afirmó definitivamente en la Organización del Conocimiento, justamente como elemento de la terminología de los tesauros.

Quienes defienden la vigencia de los procedimientos de control de vocabulario y de la construcción de vocabularios controlados, pueden sostener su posición en al menos cuatro comprobaciones: se publican regularmente versiones actualizadas de los estándares para la construcción y gestión de lenguajes de este tipo, por parte de instituciones de indiscutible respaldo y referencia internacional, como la *International Standard Organization* (ISO); estos lenguajes dan cuenta de la inserción y el uso de estos vocabularios en sistemas de información que funcionan en ambientes digitales, y promueven la interoperabilidad entre sí y con los sistemas de información a los que dan cobertura temática; surgen sistemas de organización del conocimiento que tratan de satisfacer nuevas demandas y requerimientos de usuarios de servicios en línea; se verifica la presencia de la temática en los *syllabi* de las carreras y en la producción regular de conocimiento de Bibliotecología y Ciencia de la Información.

La apelación constante de las personas al uso del lenguaje natural cuando buscan información en entornos digitales, no ha provocado una debilitación del potencial ni de la eficiencia de los vocabularios controlados. Por el contrario, es perceptible en la organización interna de taxonomías web, sitios web institucionales y portales temáticos, la preocupación por encajar los contenidos en moldes formales que aplican herramientas de normalización y de presentación de datos propios de los mecanismos tradicionales de control de vocabulario. De hecho, la organización en facetas, que es otra modalidad de organización controlada de términos, se utiliza frecuentemente en paquetes de datos disponibles en Internet.

Se han presentado en este artículo en forma sistemática nueve procedimientos de control de vocabulario, y se han identificado cinco funciones típicas. Apoyado en esas funciones y procedimientos, el control de vocabulario sigue asegurando organizaciones racionales y consistentes del conocimiento que pueden verse como interpretaciones posibles de la realidad, como estructuras formales a partir de las cuales es posible identificar las subdivisiones, los elementos y los tópicos más representativos de una disciplina o una temática.

La antigua controversia “lenguaje natural vs. vocabulario controlado” debería considerarse superada a través del reconocimiento de que en la actualidad prevalece la integración de métodos y modalidades, antes que la elección de una de las dos formas de representación y acceso. En un ambiente donde es posible ampliar las posibilidades de etiquetamiento y el acceso a la información pertinente a través de innumerables puntos de acceso, controlados o no, a los cuales puede llegarse incluso por medio de ayudas automáticas y semi-automáticas, elección implica empobrecimiento, y la inversión puesta en la innovación tecnológica y en la generación de herramientas para la recuperación, asociación y uso de información y de documentos digitales de todo tipo, va claramente en la dirección contraria.

La realidad, sea material o virtual, sigue exigiendo un ordenamiento racional para su mejor dominio y explotación desde una perspectiva humana. El control de vocabulario tiene un importante rol a jugar, como heredero de una tradición normalizadora, homogeneizadora de lo heterogéneo, ordenadora de los metadatos correspondientes a una producción documental cada vez más diversificada y numerosa.

La investigación que permita medir la fiabilidad y la eficiencia de los sistemas de información, a través del desempeño de las etiquetas temáticas controladas o semi-controladas irá marcando las señales del camino que deberá recorrerse en el futuro en esta materia.

REFERENCIAS

- ABD AZIZ, M.F. Controlled vocabulary in the digital age. *E-journal*, v. 29, n.1, p. 1-4, 2011.
- AITCHISON, J.; CLARKE, D. The thesaurus: a historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*, v. 37, n. 3/4, p. 5-21, 2004.
- AITCHISON, J.; GILCHRIST, A.; BAWDEN, D. *Thesaurus construction and use: a practical manual*. Londres: ASLIB, 2000.
- AMERICAN INSTITUTE OF CHEMICAL ENGINEERS. *Chemical Engineering Thesaurus: a wordbook for use with the concept coordination system of information storage and retrieval*. New York: AICE, 1961.
- AUSTIN, D. Vocabulary control and information technology. *Aslib Proceedings*, v. 38, n. 1, p.1-15, 1986.
- BALL, N.T. Committee on Organization of Information. *American Documentation*, v.1, n.1, p. 24-34, Jan. 1950.
- BARITÉ, M. et al. Diccionario de Organización del Conocimiento: clasificación, indización, terminología. Montevideo: PRODIC, 2013. Disponible en línea: www.prodic.edu.uy/investigacion/KOD [consultado el 15 de marzo de 2015]
- BATES, M. A tour of information science through the pages of JASIS. *Journal of the American Society for Information Science*, v. 50, n. 11, p. 975-993, 1999.
- BHATTACHARAYYA, K. The effectiveness of natural language in Science indexing and retrieval. *Journal of Documentation*, v. 30, p. 235-254, 1974.
- BEGHTOL, C. Universal concepts, cultural warrant and cultural hospitality. In: INTERNATIONAL ISKO CONFERENCE, 7., 2002, Granada. *Proceedings...* Würzburg: Ergon Verlag, 2002. p. 45-49. (Edited by M.J. López-Huertas).
- BLACK, H. An approach to a theory of subject headings. *College & Research Libraries*, v. 7, n. 3, p. 244-248, Jul. 1946.
- BRITISH STANDARDS INSTITUTION (2005-2008). Structured vocabularies for information retrieval: Guide: BS 8723-1. London: BSI. Consta de 5 partes: Parte 1: Definitions, symbols and abbreviations (2005). Parte 2: Thesauri (2005). Parte 3: Vocabularies other than thesauri (2007). Parte 4: Interoperability between vocabularies (2007). Parte 5: Formats and protocols for data exchange (2008).
- BUCHAN, R.L. Intertwining thesauri and dictionaries. *Information Services & Use*, v. 9, n. 3, p. 171-175, 1989.
- CATARINO, M.E.; BAPTISTA, A.A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. *Datagramazero*, v. 8, n. 3, 2007. Disponible en: http://dgz.org.br/jun07/Art_04.htm [consultado el 15 de marzo de 2015]

- CAVALCANTI, C. *Indexação & Tesouro: metodologia & técnicas*. Ed. preliminar. Brasília: ABDF, 1978.
- CENTELLES, M. Taxonomías para la categorización y la organización de la información en sitios web. *Hipertext.net*, 3 mayo 2005. Disponible en: <http://www.upf.edu/hipertextnet/numero-3/taxonomias.html> [consultado el 15 de marzo de 2015]
- CHU, H. *Information representation and retrieval in the digital age*. s.l.: ASIS&T, 2003.
- CINTRA, A.M.M. et al. *Para entender as linguagens documentárias*. 2a. ed. São Paulo: Polis; APB, 2002.
- CLEVERDON, C. Optimizing convenient online access to bibliographic databases. *Information Services and Use*, v. 4, n.1, p. 37-47, 1984. También publicado en WILLETT, P. (Ed). *Document Retrieval Systems*. London, UK.: Taylor Graham, 1988. p. 32-41.
- COMMITTEE ON ORGANIZATION OF INFORMATION. [Sin título]. *American Documentation*, v. 1, n. 3, p. 154-163, Aug. 1950.
- COSTELLO JR., J.C. Uniterm indexing principles, problems and solutions. *American Documentation*, v. 12, n. 1, p. 20-26, Jan. 1961.
- COWIE, A.P. Learners dictionaries in a historical and a theoretical perspective. In: HERBST, T. et al. (Ed.). *The Perfect Learners' Dictionary*. Tübingen: Niemeyer, 1999. p. 3-14.
- CURRÁS, E. *Ontologías, taxonomía y tesoros: manual de construcción y uso*. 3a. ed. actualizada y ampliada. Gijón: Trea, 2005.
- Da SILVA, F.C.C.; Da SILVA, R.C.C. Axiomas e perspectivas da evolução das linguagens documentárias na web. In: Da SILVA, F.C.C.; De SALES, R. (Org.). *Cenários da organização do conhecimento: linguagens documentárias em cena*. s.l.: Thesaurus, 2011. p. 111-128.
- DODEBEI, V.L.D. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.
- DUCROT, O.; TODOROV, T. *Diccionario enciclopédico de las ciencias del lenguaje*. 20. ed. México, D.F.: Siglo Veintiuno, 1998. 421p.
- ENGINEERS JOINT COUNCIL. Rules for preparing and updating Engineering Thesaurus. New York: EJC, 1965.
- ESTEBAN GARCÍA, M. de los M. *El C-TEST: alternativa o complemento de otras pruebas en el aprendizaje del inglés como lengua extranjera: Memoria para optar al grado de doctor: Bajo la dirección del doctor Honesto Herrera Soler*. Madrid: Universidad Complutense de Madrid, 2007.
- FAULSTICH, E. *Base metodológica para pesquisa em Socioterminologia*. Brasília: Universidade de Brasília, Departamento de Lingüística, Línguas Clássicas e Vernácula 1995.
- FIDEL, R. Who needs controlled vocabulary? *Special Libraries*, v. 83, n. 1, p. 1-9, 1992.
- FOSKETT, A.C. *The subject approach to information*. 5th ed. London: Library Association, 1996.
- FUCHS, N.E. (Ed.). *Controlled Natural Language: Workshop on Controlled Natural Language, CNL 2009*. Berlin: Springer, 2010.
- GARSHOL, L.M. Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, v. 30, n. 4, p. 378-391, 2004.
- GIL LEIVA, I. *La automatización de la indización de documentos*. Gijón: Trea, 1999.
- _____. *Manual de indización: teoría y práctica*. Gijón: Trea, 2008.
- GIL URDICIAIN, B. *Manual de lenguajes documentales*. Madrid: Noesis, 1996.
- GILCHRIST, A.; KIBBY, P. *Taxonomies for business: access and connectivity in a wired world*. London: TFPL, 2000.
- HANSSON, O. *The state of research on folksonomies in the field of Library and Information Science: A Systematic Literature Review*. Lund: Lunds Universitet. 2014. Disponible en: <http://lup.lub.lu.se/luur/load?func=downloadFile&recordOId=4451166&fileOId=4451168> [consultado el 15 de marzo de 2015]
- HUDON, M. Preparing terminological definitions for indexing and retrieval thesauri: a model. In: *ADVANCES in Knowledge Organization*. Würzburg: Ergon Verlag, 1996. v.5, p. 363-369.
- HULME, E.W. Principles of book. *Library Association Record*, v. 13, p. 444-449, 1911.
- HYSLOP, M.R. Sharing Vocabulary Control. *Special Libraries*, v. 56, p. 709, Dec. 1965.
- INSTITUTO COLOMBIANO PARA EL FOMENTO DE LA EDUCACIÓN SUPERIOR. *Lista de encabezamientos de materia para bibliotecas (LEMB)*. 2. ed. Bogotá: ICFES, Procultura, 1985.
- INTERNATIONAL STANDARD ORGANIZATION. *International Standard ISO 25.964-1: Information and Documentation: Thesauri and interoperability with other vocabularies*. Geneva: ISO, 2011. Consta de 2 partes: Parte 1: Thesauri for information retrieval. Parte 2: Interoperability with other vocabularies.
- JACKSON, P.; MOULINIER, I. *Natural language processing for online applications: text retrieval, extraction and categorization*. 2nd. ed. rev. Philadelphia: Benjamins, 2007.
- JORNA, K.; DAVIES, S. Multilingual thesauri for the modern world: no ideal solution? *Journal of Documentation*, v. 57, n. 2, p. 284-295, 2001.

- KLINGBIEL, P.H. *The future of indexing and retrieval vocabularies*. Alexandria: Defense Documentation Center, 1970.
- KWASNIK, B. The role of classification in knowledge representation and discovery. *Library Trends*, v. 48, n. 1, p. 22-47, 1999.
- LANCASTER, F.W. *Vocabulary control for information retrieval*. Washington: Information Resources Press, 1972.
- _____. Vocabulary control in information retrieval systems. In: *ADVANCES in Librarianship*. London: Academic Press, 1977. v. 7, p. 1-40. Edited by Melvin Voight and Michael Harris.
- _____. *Vocabulary control for information retrieval*. 2nd. ed. Arlington: IRP, 1986. 270 p.
- LEATHERDALE, D. *Categorías de materias AGRIS, AGRINTER*. Costa Rica, 1980.
- LEE, J.M. E. Wyndham Hulme: a reconsideration. In: RAYWARD, W.B. (Ed.). *The variety of Librarianship: essays in honour of John Wallace Metcalfe*. Sydney: LAA, 1976.
- MANIEZ, J. *Los lenguajes documentales y de clasificación*. Salamanca: Fundación Germán Sánchez Ruipérez, 1993.
- MARCONDES, C.H. O papel das relações semânticas na organização e representação do conhecimento em ambientes digitais. In: Da SILVA, F.C.C.; De SALES, R. (Org.). *Cenários da organização do conhecimento: linguagens documentárias em cena*. s.l.: Thesaurus, 2011. p. 129-168.
- MARTÍNEZ, A.M.; VALDEZ, J.C. Indización y clasificación en bibliotecas. Buenos Aires: Alfagrama, 2008.
- MCILLWAIN, I.C. Some problems of context and terminology. *Knowledge Organization*, v. 25, n. 4, p. 144-146, 1998.
- MÉNARD, E. Images: indexing for accessibility in a multi-lingual environment – challenges and perspectives. *The Indexer*, v. 27, n. 2, p. 70-76, 2009.
- MOREIRO, J.A. Documentación y lingüística: conceptos de relación esenciales. *Ciencias de la Información*, v. 25, n. 4, p. 202-211, 1994.
- MORRIS, I. Principles of vocabulary Control: 4). English. *Language Teaching*, v. 2, n. 3, p. 57-62, Nov. 1947.
- MOTE, L.J. Reasons for the variations in the information needs of scientists. *Journal of Documentation*, v. 18, n. 4, p. 169-175, 1962.
- NATIONAL INFORMATION STANDARDS ORGANIZATION. Guidelines for the construction, format and management of monolingual controlled vocabularies : an American National Standard ANSI/NISO Z39.19-2005(R2010) developed by the National Information Standards Organization. Bethesda: NISO Press, 2010. Disponible en: http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf [consultado el 15 de marzo de 2015]
- NAUMIS, C. *Los tesauros documentales y su aplicación en la información impresa, digital y multimedia*. México: UNAM; Buenos Aires: Alfagrama, 2007.
- NORUZI, A. Folksonomies: (Un)controlled vocabulary? *Knowledge Organization*, v. 33, n. 4, p. 199-203, 2006.
- OGDEN, C.K. *Basic English: a general introduction with rules and grammar*. London: Paul Treber, 1930.
- PALMER, H. The history and present state of the movement towards vocabulary control: Part 1. *Bulletin of the Institute for Research in English Teaching*, v. 120, p. 14-17, Jan. 1936a.
- _____. The history and present state of the movement towards vocabulary control: Part 2. *Bulletin of the Institute for Research in English Teaching*, v. 121, p. 19-23, Feb. 1936b.
- _____. *A grammar of english words*. London: Longmans, 1938.
- PETERSON, E. Beneath the Metadata: some philosophical problems with Folksonomy. *D-Lib Magazine*, v. 12, n. 11, 2006. Disponible en: <http://www.dlib.org/dlib/november06/peterson/11peterson.html> [consultado el 15 de marzo de 2015]
- RAO, M. Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics. *Journal of the American Society for Information Science*, v. 49, n. 10, p. 881-887, 1998.
- ROBERTS, N. The pre-history of the information retrieval thesaurus. *Journal of Documentation*, v. 40, n. 4, p. 271-285, Dec. 1984.
- ROE, S.K.; THOMAS, A.R. The thesaurus: review, renaissance and revision. Binghamton, N.Y.: Haworth, 2004.
- SACHS, M.; SMIRAGLIA, R. From encyclopedism to domain-based ontology of knowledge management: the evolution of the Sachs Classification (SC). In: INTERNATIONAL ISKO CONFERENCE, 8., 2004. London. *Proceedings...* Würzburg: Ergon Verlag, 2004. p. 167-172. Edited by I. C. McIlwaine.
- SALTON, G. A new horizon for Information Science: Letter to the editor. *Journal of the American Society for Information Science*, v. 47, n. 4, p. 333, 1996.
- SALVADOR, G. *Semántica y lexicología del español: estudios y lecciones*. Madrid: Paraninfo, 1985.

- SÁNCHEZ-CUADRADO, S.; COLMENERO-RUIZ, M.J.; MOREIRO, J.A. Tesoros: estándares y recomendaciones. *El Profesional de la Información*, v. 21, n. 3, p. 229-235, 2012.
- SHAPIRO, F.R. Contributions to the history of library terminology. *Library Quarterly*, v. 59, n. 2, p. 95-115, Apr. 1989.
- SHIELDS, M. Subject headings in Physics. *College & Research Libraries*, v. 6, n. 2, p. 185-187, Mar. 1945.
- SIMÓN, L. et al. Estructuras sintagmáticas en tesauros: aproximación cualitativo-cuantitativa. In: JORNADAS DE INTERCAMBIOS Y REFLEXIONES ACERCA DE LA INVESTIGACIÓN EN BIBLIOTECOLOGÍA, 3., 2013, La Plata, Argentina. *Actas...* La Plata: Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata, 2013. p. 1-15. Disponible en: http://www.jornadabibliotecologia.fahce.unlp.edu.ar/jornadas-2013/actas-2013/simon_fontans_barite_duarte_.pdf [consultado el 15 de marzo de 2015]
- SISTEMA de Clasificación Decimal: diseñado originalmente por Melvil Dewey. Traducción de la ed. 20 en inglés. Santa Fe de Bogotá: Rojas Eberhard, 1995. 4 vol.
- SLAVIC, A. On the nature and typology of documentary classifications and their use in a networked environment. *El profesional de la información*, v. 16, n. 6, p. 580-589, nov.- dic. 2007.
- STAMBUK, A. Metaphor in scientific communication. *Meta*, v. 43, n. 3, p. 373-379, 1998.
- STONE, A. The LCSH century: a brief history of the library of congress subject headings, and introduction to the centennial essays. *Cataloging & Classification Quarterly*, v. 29, n. 1/2, 2000. Disponible en: <http://www.catalogingandclassificationquarterly.com/ccq29nr1-2ed.htm> [consultado el 15 de marzo de 2015]
- SVENONIUS, E. Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science*, v. 37, n. 5, p. 331-340, 1986.
- TAUBE, M. *Uniterm System of Indexing Operating Manual*. Washington: Documentation Inc., 1955.
- THORNDIKE, E.L.; LORGE, I. *The teacher's word book of 30,000 words*. Nueva York: Teachers College Columbia University, 1944.
- U.S. Defense Documentation Center. *Thesaurus of Astia Descriptors*. Arlington, Va., 1962.
- U.S. National Institutes of Health. *Medical and health related sciences thesaurus*. Washington: Department of Health, Education, and Welfare, Public Health Service, 1963.
- VARELA BARRAZA, J.A. et al. Las 5000 palabras más frecuentes en los libros de texto oficiales de la educación básica en México. *Revista Electrónica de Investigación Educativa*, v. 15, n. 3, ene. 2013. Disponible en: http://www.scielo.org.mx/scielo.php?pid=S1607-40412013000300008&script=sci_arttext [consultado el 15 de marzo de 2015]
- WAY, E.C. *Knowledge representation and metaphor*. Oxford: Intellect, 1991.
- WEST, M. *A general service list of english words*. London: Longman, Green and Co., 1953.