

---

# El control de vocabulario en la era digital: revisión conceptual

---

*Vocabulary control in the digital age: a conceptual review*

---

**Mario BARITÉ**

Facultad de Información y Comunicación, Universidad de la República, Uruguay  
mbarite@gmail.com

## Resumen

Revisión conceptual de la evolución del control de vocabulario hasta el momento actual, la era digital. Se establecen cuatro eras en la historia de la representación y la recuperación de información, en relación con la difícil convivencia entre las técnicas de lenguaje natural y las de lenguaje controlado. Se hace una breve historia del término 'control de vocabulario' mostrando cómo el mismo se originó en el ámbito de la enseñanza del inglés como una segunda lengua, a comienzos de la década de 1930. Se desarrollan a continuación los cinco objetivos del control de vocabulario: restricción, desambiguación, normalización, jerarquización y relación. Se analiza la situación del control de vocabulario en un entorno digital regido por usuarios autosuficientes que utilizan preferentemente el lenguaje natural, en tanto los responsables de portales y sitios web formalizan sus datos siguiendo criterios tradicionales de normalización.

**Palabras clave:** Control de vocabulario. Vocabularios controlados. Representación del conocimiento. Organización del conocimiento. Era digital.

## 1. Introducción

El control de vocabulario es un conjunto de técnicas y procedimientos que se ejercen sobre expresiones lingüísticas denominadas 'términos', que representan conceptos en un determinado contexto temático. En Organización del Conocimiento (OC) estas técnicas y procedimientos se utilizan en el proceso de creación, desarrollo o revisión de sistemas de organización del conocimiento, para fines de indización de documentos y recursos de información. En este ámbito, el control de vocabulario permite establecer formas normalizadas de los descriptores o encabezamientos de materia seleccionados para integrar el repertorio terminológico respectivo (tesauros, listas, taxonomías). Asimismo promueve la representación de relaciones estables entre conceptos, las cuales van reconstruyendo por acumulación y vinculación la estructura conceptual de un dominio especializado. Desde esta perspectiva, el control de vocabulario puede ser visto como una modalidad de normalización de la terminología utiliza-

## Abstract

Conceptual review of the history of vocabulary control, up to the present time, the digital age. Four different periods are established in information representation and retrieval in relation to the difficult co-existence between natural and controlled language techniques. A brief history of the term "vocabulary control" is traced showing how it originated in the teaching of English as a Second Language in the early 1930s. The five objectives of vocabulary control i.e. restriction, disambiguation, standardization, hierarchy and relation are developed. The author performs an analysis on the situation of vocabulary control in a digital environment, ruled by self-sufficient users who prefer natural language, while the people in charge of Internet portals and websites fixes their data following traditional standardization criteria.

**Keywords:** Vocabulary control. Controlled vocabularies. Knowledge representation. Knowledge organization. Digital age.

da en sistemas de información, para favorecer los procesos de búsqueda, acceso y recuperación temática de información.

En este trabajo se plantea el siguiente derrotero: en primera instancia se analizan cuestiones vinculadas con la pertinencia del control de vocabulario, al tiempo que se establecen las distintas fases del debate todavía inacabado entre el uso del lenguaje natural y de los vocabularios controlados como herramientas de representación temática del conocimiento. En segundo lugar, se propone una probable historia del término "control de vocabulario", que documenta su origen en el área de la enseñanza del idioma inglés como segunda lengua, y brinda datos acerca de su incorporación a la terminología de la Organización del Conocimiento. En tercer término, se desarrollan los objetivos del control de vocabulario, a partir de la revisión de literatura realizada y del análisis comparativo de estándares y manuales de referencia en la temática. A continuación se analiza el nuevo escenario planteado por la web semántica y el poten-

cial alternativo de estandarización de nuevos sistemas de organización del conocimiento tales como las folksonomías y las ontologías. Por último, se establecen conclusiones a partir del estudio realizado.

## 2. Pertinencia del control de vocabulario

El control de vocabulario supone la preexistencia de un vocabulario, seleccionado previamente. Así, el control de vocabulario no incide sobre la calidad de la selección previa, pero necesita disponer de una caja terminológica suficiente, pertinente y calificada.

Los estándares desarrollados a escala internacional (British Standards Institution, 2005; National Information Standards Organization, 2005; International Standard Organization, 2011), juegan un papel trascendente: proponen reglas de aplicación universal y buscan unificar criterios metodológicos para desarrollar vocabularios controlados, clasificar e indizar recursos de información convencionales o electrónicos.

Para los defensores del control de vocabulario, la normalización opera favorablemente tanto sobre la indización de documentos (aportando una estructura de referencia para el indizador), como sobre la recuperación de información. Subyace la idea de que tales operaciones jerarquizan y mejoran la calidad de los términos de indización y las búsquedas temáticas, en contraposición al uso del lenguaje natural, no controlado ni estructurado; idea interpelada cotidianamente por la mejora continua de las aplicaciones informáticas basadas en el lenguaje.

Existe una relación de estrecha interdependencia entre los conceptos de control de vocabulario y vocabulario controlado. En efecto, un vocabulario controlado solo puede construirse a partir de sucesivas operaciones de control de vocabulario. Inversamente, el control de vocabulario —aunque puede utilizarse en otros ámbitos como la creación de diccionarios terminológicos— adquiere sentido pleno si se realiza en el contexto del desarrollo de vocabularios controlados.

En un intento de sistematización de abordajes teórico-metodológicos en Organización del Conocimiento, Hjørland (2008) identificó seis diferentes enfoques o aproximaciones: 1) un abordaje tradicional, difícil de definir porque no cuenta con una teoría unificada que corresponda a ese concepto, sino con diferentes prácticas y algunas teorías aisladas de cómo organizar el conocimiento; 2) la corriente analítico sintética o facetada de Ranganathan; 3) la línea de recuperación de información (*information retrieval*);

4) enfoques cognitivos y orientados al usuario; 5) abordajes bibliométricos; y 6) el análisis de dominio (*domain analysis*).

Entre los aportes que Hjørland (2008, p. 89) situó en relación con el abordaje tradicional de la Organización del Conocimiento menciona al principio de vocabulario controlado, junto a la Regla de Cutter sobre la especificidad, el principio de garantía literaria de Hulme, y el principio de organización desde lo general hacia lo específico. La escueta mención que realiza sobre el principio de vocabulario controlado dice que este “es esencialmente una forma de evitar los sinónimos y homónimos como términos de indización utilizando un vocabulario estandarizado.”

Pese al transcurrir de las décadas, las cuestiones de control de vocabulario siguen siendo vistas como nucleares y altamente significativas en la formación universitaria en Bibliotecología, Documentación y Ciencia de la Información, tanto a nivel de grado como de posgrado, como surge de cualquier análisis sumario de programas de estudio y *syllabi* de universidades de todos los continentes en los que se dictan esas carreras, como por ejemplo, la Indiana University School of Library and Information Science de Estados Unidos, la Universidad de La Plata, en Argentina, o la Universidad Complutense de Madrid en España.

La misma preocupación y vigencia se expresa en congresos, simposios y encuentros de la especialidad a lo largo y ancho del mundo, y si bien no hay estudios fehacientes al respecto, se percibe que el interés por el control de vocabulario no parece decaer en la producción científica y especializada del área. Por otra parte, los estándares de mayor influencia están elaborados sobre la base del control de vocabulario en todas sus modalidades y expresiones, en consonancia con el hecho de establecer sus normas para “vocabularios estructurados” o “vocabularios controlados” (British Standard Institution, 2005; National Information Standards Organization, 2010; International Standard Organization, 2011).

Chu (2003) establece cuatro eras en la historia de la representación y la recuperación de información, que se desarrollan aquí con el agregado de fechas tentativas de apogeo y otras apreciaciones de contexto:

(1) *Primera mitad del siglo veinte*: Aplicación preferente del lenguaje natural ante la existencia de lenguajes controlados primitivos, tales como las primeras listas de encabezamientos. A raíz de la aplicación del lenguaje natural para indizar documentos, empezaron a identificarse los problemas planteados por los homógrafos y los

sinónimos, y se comenzó a comprender que la utilización exclusiva del lenguaje natural provocaba problemas de recuperación de información. Puede afirmarse que Taube con su formulación del sistema Unitérmino y su propuesta de indexación coordinada (*coordinate indexing*) incidió en esta discusión y cerró esta primera etapa (Taube, 1955; Bates, 1999; Mesquita, Souza e Porto, 2014).

(2) *Segunda mitad del siglo veinte*: Introducción de vocabularios con procedimientos de control que se van mejorando y sistematizando: mayor desarrollo de listas de encabezamientos de materia y de sistemas de clasificación precoordinados, y evolución paulatina de la metodología de construcción de tesauros y lenguajes similares. En este período se verificó una tendencia creciente a favor del control de vocabulario, y se inició al mismo tiempo una discusión que no se ha resuelto hasta el presente, entre partidarios del lenguaje natural y del control de vocabulario (Rao, 1998; Dodebei, 2002).

(3) *Desde 1990 en adelante*: Resurgimiento del lenguaje natural como consecuencia de las nuevas aplicaciones informáticas de recuperación por palabras clave. El control de vocabulario sigue utilizándose en forma predominante en sistemas de información tradicionales, en tanto las técnicas de recuperación por lenguaje natural van de la mano con el desarrollo de las primeras aplicaciones informáticas destinadas a explorar el contenido de bases y bancos de documentos a texto completo. El surgimiento de Internet provoca el desarrollo de herramientas de recuperación basadas en el lenguaje natural, y la profundización de los estudios de *Natural Language Processing* o NLP (Jackson & Mouligner, 2007).

(4) *Desde 2000 en adelante*: Surgimiento de la web 2.0 asociada al desarrollo de modelos colaborativos, centrados en los intereses y la acción de los usuarios, a través de interfaces evolucionadas, que se apoyan en el lenguaje natural para la representación y recuperación de información. Chu señala que si bien las fronteras entre una era y otra no admiten cortes claros, a su criterio las dos primeras fases ya fueron superadas históricamente, en tanto que la tercera y la cuarta permanecen y evolucionan con cierto paralelismo (Chu, 2003).

En el mismo trabajo Chu afirma que las razones por las que coexistieron las dos tendencias durante tanto tiempo, tienen que ver con los pro y los contra de cada una de las dos vertientes de representación y recuperación de información, en relación a los siguientes siete tópicos: a) los *sinónimos* y su manejo; b) los *homógrafos* y la

ambigüedad por falta de contexto provocada por la polisemia; c) las cuestiones de *sintaxis*; d) el problema de la *precisión* (*accuracy*), debido entre otras razones a la falta de especificidad de los vocabularios controlados, y a las dificultades de ajuste entre la connotación y la denotación de cada término en relación a diferentes categorías de usuarios, y a la misma imprecisión del lenguaje; e) la cuestión crucial de la *actualización*, ya que a diferencia de los vocabularios controlados, el lenguaje natural no requiere ningún proceso específico; f) el *costo*, pues en principio el trabajo con lenguaje natural es menos caro porque no requiere entrenamiento ni mantenimiento, aunque las dificultades que genera al momento de la recuperación implican pérdida de tiempo para los usuarios, y por lo tanto un costo que esto puede implicar en diferentes situaciones; y, g) la *compatibilidad*, porque para los sistemas basados en el lenguaje natural, el emparejamiento (*switching*) o la migración de un sistema a otro es más sencillo y directo, porque el lenguaje natural siempre es compatible consigo mismo (Chu, 2003, p. 52ss.), si bien exhibe las inconsistencias propias de la relación natural de los sinónimos y los términos polisémicos entre sí.

### 3. Breve historia del término “control de vocabulario”

¿Desde cuándo se habla de control de vocabulario en el ámbito de la Organización del Conocimiento? No hay para esta pregunta una respuesta precisa y ajustada. No obstante, es posible aseverar que casi seguramente el término fue extrapolado desde el área del vocabulario básico para el aprendizaje y el estudio de las lenguas hacia la Bibliotecología y la Ciencia de la Información. En efecto, hacia 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 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). Pal-

mer 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), dando primacía al criterio de frecuencia de las palabras en la lengua.

En forma paralela, hubo ya en la literatura de la Bibliotecología y la Ciencia de la Información en la década de los 40, varias aproximaciones a la utilización de la palabra “control” en forma aislada, pero expresiva. Puede interpretarse que la palabra “control” se enmarca en cuestiones generales de planificación, gestión y administración, y que por lo tanto guarda relación con procedimientos racionales organizados que pretenden ser efectivos en el manejo de un segmento de la realidad. Controlar implicar incidir desde una posición de gobierno, dominio o conocimiento, 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.”

En el mismo amanecer de la década de 1950 comienza a aparecer de forma titubeante en distintos artículos, junto con la palabra *control* el término *vocabulary control*. Salvo mejor prueba, la primera cita hallada por este autor, aparece en el primer número de la revista *American Documentation*, en un informe de un Comité sobre Organización de la Información (Ball, 1950). Queda por descubrir el eslabón perdido que permitió que el término pasara del ámbito de la enseñanza del inglés como una segunda lengua, al de la Bibliotecología y Ciencia de la Información. La conexión será descubierta pronto porque como enseña Hulme a través de su principio de garantía literaria, toda la expresión del conocimiento humano se expresa en la documentación, y por tanto solo falta dar con los documentos que hagan visible ese nexo todavía no encontrado.

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 el área (Comitee on Organization of Infomation, 1950).

El abstract de un documento del año 1961, tomado también 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). Este ejemplo tomado al azar, da 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. El uso del término se fue expandiendo luego en otras direcciones dentro del área, y así fue posible dar con expresiones tales como “control de búsqueda”, “control de autoridades” o “control bibliográfico universal”, que siguen vigentes y generan espacios de desarrollo para la investigación y el trabajo profesional.

En lo que hace a la Organización del Conocimiento, la palabra “control” siguió 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.

#### 4. Objetivos del control de vocabulario

A partir del la revisión de literatura y del análisis comparado de los estándares de mayor prestigio en el mundo (British Standards Institution, 2005; National Information Standards Organization, 2010; International Standard Organization, 2011), puede afirmarse que el control de vocabulario cumple con cinco objetivos diferentes y complementarios:

1. *Restricción*. El control de vocabulario establece límites en cuatro direcciones, a saber:

i) La cantidad de términos a ser incluidos en un vocabulario controlado. En esta dirección se pretende dar la mayor cobertura posible de un área del conocimiento con la menor cantidad de términos significativos. 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. La extensión de un descriptor afecta siempre 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).

iii) El alcance conceptual de cada término. Se tiende a asignar a cada uno un solo significado, en cumplimiento del principio de univocidad (un término un significado, y viceversa). 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 agrega en esa fuente 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 tesoro 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 trata de limitar 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 del término en la estructura conceptual de referencia. Desde este punto de vista, 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.

Ya se ha mencionado que las ambigüedades provocadas por las formas sinonímicas y las variantes se esclarecen a través de relaciones de equivalencia. Según Bakhtin (Bakhtin apud Cintra *et al.*, 2002, p. 67-68),

[...] en el plano ideológico, la palabra es una unidad neutra, apta para adecuarse a diferentes patrones 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.

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, por el cual 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.

iii) Indicación de notas de alcance. Estas notas son usadas (National Information Standards Organization, 2010, p. 22),

[...] 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.

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 equivalencia, jerárquicas y asociativas.

iv) Definición de cada término. Constituye la forma extrema y 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). Desde la óptica de la desambiguación, el control de vocabulario se constituye

en un proceso de determinación conceptual y terminológica.

**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. De hecho, podría considerarse que el control de vocabulario es la vertiente que corresponde al control 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). La normalización se alcanza mediante reglas formales para la elección de singulares y plurales, ortografías correctas, formas gramaticales preferidas y cuestiones similares.

**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. Esta jerarquización debería realizarse a partir de la garantía utilizada. Así, por ejemplo, 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 uso, deberían tomarse en cuenta las expresiones mayormente utilizadas por los usuarios 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 dentro 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 permite establecer criterios uniformes y racionales para la selección y jerarquización de terminología. Cabe agregar que el estándar norteamericano propone el respaldo en alguna de las tres garantías que menciona

(literaria, de usuario, organizacional). Asimismo, identifica una forma de control de vocabulario no jerárquica: el anillo de sinónimos. En este caso, el conjunto de formas sinonímicas identificadas como elementos de indización o de recuperación para un concepto, son exhibidas todas en el mismo rango de jerarquía, pues todas ellas son bocas de acceso temático a la base de datos (National Information Standards Organization, 2010).

**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. Si bien las tipologías tradicionales de relaciones intra y entre términos han dado firmeza al proceso de control de vocabulario, deben considerarse estrategias diferentes, pues las disciplinas tienen distintas consistencias terminológicas, conforme con la estabilidad de las relaciones internas de sus conceptos (Mote, 1962; Svenonius, 1986). La tipología de relaciones está restringida a un número cerrado de categorías, dentro de las cuales se ubican todas las posibilidades existentes de vinculación paradigmática o sintagmática entre objetos y/o conceptos. Las relaciones entre términos reguladas por los estándares y establecidas en los manuales más relevantes (Aitchison, Gilchrist & Bawden, 2000; Currás, 2005) se han mantenido estables a lo largo del tiempo y son:

a) Relaciones de equivalencia, que comprenden las de sinonimia, cuasi sinonimia, variantes lexicales, algunas relaciones de jerarquía que 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").

b) Relaciones jerárquicas, que comprenden las relaciones genéricas (también llamadas de género/especie); las partitivas o todo/parte; las relaciones de casos particulares (también llamadas enumerativas) como en el ejemplo "Montañas/Aconcagua", y las relaciones polijerárquicas.

c) Relaciones asociativas, que comprenden las relaciones entre disciplinas y sus objetos de estudio, entre procesos u operaciones y sus agentes, entre ocupaciones y personas que las ejercen, entre acciones y el producto de esas

acciones, entre acciones y propiedades asociadas con ellas, entre acciones y sus sujetos pacientes, entre entidades y sus propiedades, relaciones genéticas, de causa-efecto, las establecidas entre materias primas y productos, relaciones funcionales y relaciones semánticas.

d) Relaciones personalizadas, las que expresan una vinculación que requiere un tratamiento particular. La norma ISO 25.964-1 ejemplifica con la relación entre una sigla y el nombre desarrollado de una institución (International Standard Organization, 2011, p. 67).

Los estándares en curso colocan diferentes énfasis respecto a los objetivos que persigue el control de vocabulario. 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).

## 5. Control y entorno digital: nuevas herramientas

Son objeto de interrogación en forma recurrente en la literatura, cuestiones tales como ¿el control de vocabulario sigue siendo necesario en un entorno de información claramente orientado a la evolución de la tecnología digital? ¿Siguen proveyendo los vocabularios controlados mejor acceso a la información que el lenguaje natural? ¿No es hora de adaptar las cuestiones de organización temática al perfil de los nuevos usuarios, en especial los nacidos digitales? ¿Cómo justificar el desarrollo de nuevos vocabularios controlados? (Abd Aziz, 2011). Hay todavía una cuestión más sustancial: con la evolución de las herramientas y aplicaciones informáticas, ¿no ha llegado el momento de superar la dicotomía, que puede resultar falsa, entre lenguaje natural y lenguaje controlado?

Cada una de las preguntas formuladas podría llenar un estante digital completo de libros y artículos científicos soportados en formatos electrónicos. De hecho se escribe mucho sobre todas estas cuestiones, y la ambición que se persigue en este trabajo es dar apenas los li-

neamientos principales de la cuestión, desde la perspectiva del control de vocabulario.

En términos generales, puede decirse que, como señalan Da Silva y Da Silva (2011, p. 114),

[...] hoy estamos habituados al montaje de una infraestructura informacional que evoluciona rápidamente para la omnipresencia permanente; sin embargo, mientras esto todavía no ocurre, podemos observar que en muchos sistemas la información mantiene un control de autoridad casi dictatorial sobre el acceso. Pero el mundo digital está por derrumbar (o repaginar) todo lo que conocemos...

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 alcanzó la coexistencia de formas lingüísticas de recuperación (escribiendo en una ventana la ecuación de búsqueda) con otras basadas en el clickeo. 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 (2011, p. 141), 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."

En este marco, dos fenómenos concurrentes inciden en la evolución del debate entre lenguaje natural y control de vocabulario. El primero de ellos es el desarrollo de nuevos sistemas de organización del conocimiento (manifiestamente las ontologías y las folksonomías) cuya estructuración interna pone en cuestión el modelo tradicional de creación de vocabularios controlados.

Las ontologías y las folksonomías han demostrado ser alternativas con un gran potencial para la organización del conocimiento. Las ontologías son diseños de estructuras funcionales, que contienen entidades o elementos que se relacionan entre sí, para llevar a cabo determinados propósitos o para cumplir con ciertos objetivos, en un entorno electrónico. Son funcionales porque no siempre pretenden representar semánticamente un segmento del conocimiento o un área de actividad, sino sobre todo desplegar una red de asuntos o acciones con sus relaciones, volviendo explícitos los circuitos que en su conjunto configuran un dominio. Más que una estructura de conocimiento, una ontología es ante todo un sistema relacional de acciones que persigue tanto una gestión corporativa de calidad como la satisfacción plena del usuario. Na-

da impide, naturalmente, que una ontología se organice en torno a tópicos temáticos y que se asimile a la estructura tradicional de un tesoro, pero sus diseñadores siempre tendrán en consideración el factor funcional antes que el factor semántico.

Si bien Vickery (1997) acepta el hecho de que toda ontología es una "conceptualización" que "implica una visión simplificada del universo que se pretende representar para un propósito determinado" (Gruber, 1992), y por tanto se da una obvía "analogía con clasificaciones bibliográficas y tesauros" (Vickery, 1997, p. 284). Este autor indica que son igualmente obvias las diferencias entre ambas conceptualizaciones, en virtud de que tienen ámbitos de aplicación diferentes. Una ontología bien diseñada toma en cuenta, antes que nada y sobre todo, los intereses y los comportamientos predecibles de los usuarios, así como los objetivos concretos que estos esperan cumplir dentro de la lógica de ese dominio. Desde esa concepción, no encuadran necesariamente en las tipologías de relación (de equivalencia, jerárquicas, asociativas) propugnadas tradicionalmente por el control de vocabulario.

Por su parte, las folksonomías, también llamadas clasificaciones sociales, proporcionan una forma de indización de recursos disponibles en Internet, mediante la asignación de etiquetas (tags, labels, palabras claves), generadas por los mismos usuarios, en forma individual en la práctica, y en forma asociativa por acumulación, para categorizar contenidos tales como recursos digitales de información, fotografías o videos en línea o enlaces (Noruzzi, 2006). Los sitios y sistemas de información web así organizados tienden a ser muy populares (por ejemplo, YouTube). Señalan Da Silva y Da Silva con efusividad que "nuevos recursos tales como las folksonomías, permiten que las personas clasifiquen las informaciones conforme a sus convicciones e intereses. ¡Es la sabiduría de las multitudes en acción!" (Da Silva y 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). Visto así, todo haría suponer que las folksonomías solamente dan forma a estrategias de recuperación de información en Internet mediante el lenguaje natural proporcionado por los mismos usuarios. Sin embargo, existen ayudas que pueden ser vistas como modos de control de vocabulario: por ejemplo, el menú de opciones que se abre a medida que se van digitando letras en la ventana (una aplicación 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 una listado alfabético de entradas temáticas, tal como se acumulaba en los viejos catálogos manuales que recogían las entradas secundarias por tema. Es posible que muchas de estas clasificaciones sociales sean moderadas por analistas que eliminan errores ortográficos o inconsistencias gramaticales. De ese modo, a lo largo del tiempo estas folksonomías van constituyéndose entonces en vocabularios semicontrolados.

El segundo fenómeno que se ha generalizado en los últimos años es el desarrollo de herramientas y aplicaciones que operan como formas nuevas de 'control' lingüístico. Además de los menús que se abren a medida que se escriben letras en las ventanas de búsqueda, aparecen sugerencias cuando hay faltas de ortografía 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 (Da Silva y Da Silva, 2011, p. 123),

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, etc. Al mismo tiempo en que las empresas innovan, la búsqueda de la diferenciación crea diferentes visiones sobre como la información es tratada.

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. En la peor de las circunstancias siempre pueden sentirse confortados por los poderosos motores de búsqueda existentes, a través de los cuales cada palabra, cada signo, cada número puede constituirse en un punto de acceso.

Aún en este renovado y dinámico escenario, las operaciones de control de vocabulario y los vocabularios controlados tienen un importante



rol a jugar, como herederos de una tradición normalizadora, homogeneizadores de lo heterogéneo, ordenadores de la producción documental en cualquier formato, compiladores y sistematizadores de la terminología de los campos especializados, jerarquizadores de conceptos y conocimientos. Como establece Slavic (2007, p. 580-581),

[...] 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.

Tal vez sea el momento de aceptar que eso es lo está sucediendo y que la vieja controversia “lenguaje natural vs. vocabulario controlado” puede ser superada reconociendo que hoy día prevalece la integración de métodos y modalidades antes que la elección de una de las dos formas de acceso. En este ámbito, 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.

Por otra parte, tradicionalmente los vocabularios controlados se han nutrido del lenguaje natural. El nuevo desafío se sitúa ahora en la articulación de los probados y aprobados procesos de control de vocabulario con las nuevas aplicaciones e interfaces, y con las facilidades que ofrece la tecnología en el acceso y la organización de la información para lograr una contribución efectiva en lo que siempre ha sido su más reconocida fortaleza: la presentación racional y organizada de artefactos, conceptos y conocimientos fiables, actuales y pertinentes a los intereses de cada comunidad de usuarios.

## 6. Conclusiones

El control de vocabulario permanece como herramienta intelectual para organizar y estructurar listas de términos destinados a la representación del conocimiento a través de formas o códigos de clasificación normalizados. Permite una aproximación racional y estructurada al conocimiento a través de los sistemas de organización del conocimiento tradicionales, al tiempo que ofrece una perspectiva de usuario, favoreciendo la recuperación temática en sistemas de información, y reduciendo la ambigüedad propia del lenguaje natural. Proporciona consistencia en la descripción temática de recursos de información, aumenta la eficiencia de las herramientas de búsqueda, y la efectividad de los

sistemas de almacenamiento y recuperación de información, los sistemas de navegación en Internet y otros entornos. Hay un reconocimiento implícito de su vigencia en la continuidad de los estudios universitarios, la investigación y la producción de estándares a su respecto.

La apelación constante de las personas al uso del lenguaje natural cuando buscan información en entornos digitales, y el desarrollo de sistemas de organización del conocimiento contruidos desde novedosas perspectivas, tales como las folksonomías y las ontologías, no han 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 muy cercanos a los mecanismos tradicionales de control de vocabulario.

La investigación relativa a esta temática se ha dado en general como apoyo al desarrollo de estándares. Por ello, se hace sentir en la literatura la necesidad de orientar nuevos estudios hacia la incidencia y el aporte que el control de vocabulario puede tener y realizar para el mejor desarrollo de una cultura digital sostenida en grandes bancos y almacenes de información, y en una utilización habitual del lenguaje natural para la búsqueda y el apoyo de ayudas semi-controladas para la orientación.

Así como la expresión “control de vocabulario” surgió en el ámbito de la enseñanza del inglés como una segunda lengua, quizás haría justicia a su origen pensar hoy que su contribución teórica, de métodos y aplicaciones puede ayudar a organizar la gran antibabel que representa Internet, en tanto lengua común de una cultura digital creciente en la que todos somos, sin embargo y a la vez, tanto ciudadanos como extranjeros.

## Referencias

- Abd Aziz, Muhamad Faizal (2011). Controlled Vocabulary in the Digital Age. // E-journal. 29:1, 1-4
- Aitchison, J.; Clarke. D. (2004). The thesaurus: a historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*. 37:3/4, 5-21.
- Aitchison, J.; Gilchrist, A.; Bawden, D. (2000). *Thesaurus construction and use: a practical manual*. Londres: Aslib.
- Ball, Norman T. (1950). Committee on Organization of Information. // *American Documentation*. 1:1 (January 1950), 24-34.
- Bates, Marcia (1999). A tour of information science through the pages of JASIS. // *Journal of the American Society for Information Science*. 50:11, 975-993.
- British Standards Institution (2005). *Structured vocabularies for information retrieval: BS 8723-1*. S.I.: BSI.

- Buchan, R. L. (1989). Intertwining thesauri and dictionaries. *Information Services & Use*. 9:3, 171-175.
- Catarino, M. E.; Baptista, A. A. (2007). Folksonomia: um novo conceito para a organização dos recursos digitais na Web. // *Datagramazero*. 8:3. [http://dgz.org.br/jun07/Art\\_04.htm](http://dgz.org.br/jun07/Art_04.htm) (30-05-0214).
- Chu, Heting (2003). Information Representation and Retrieval in the Digital Age. *ASIS&T*.
- Cintra, Anna Maria Marques [et al] (2002). Para entender as linguagens documentárias. 2a. ed. São Paulo: Polis; APB.
- Committee on Organization of Information (1950). [Sin título]. // *American Documentation*. 1:3 (August 1950), 154-163.
- Costello Jr., J.C. (1961). Uniterm indexing principles, problems and solutions. *American Documentation*. 12:1 (January 1961), 20-26.
- Cowie, Anthony P. (1999). Learners dictionaries in a historical and a theoretical perspective. // Herbst, Thomas (editor) et al. *The Perfect Learners' Dictionary*. Tübingen: Niemeyer. 3-14.
- Currás, Emilia (2005). Ontologías, taxonomía y tesauros: manual de construcción y uso. 3a. ed. actualizada y ampliada. Gijón: Trea.
- Da Silva, Fabiano Couto Corrêa; Da Silva, Rodrigo Couto Corrêa (2011). Axiomas e perspectivas da evolução das linguagens documentárias na web. // Da Silva, Fabiano Couto Corrêa; De Sales, Rodrigo, organizadores. *Cenários da organização do conhecimento: linguagens documentárias em cena*. Brasil: Thesaurus. 111-128.
- Dodebei, Vera Lucia Doyle (2002). Tesauro: Linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência.
- Esteban García, María de los Milagros (2007). 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.
- Gruber, T. R. (1992). What is an ontology?. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (30-05-0214).
- Hjørland, Birger (2008). What is Knowledge Organization (KO)?. // *Knowledge Organization*. 35:2/3, 86-101.
- Hudon, M. (1996). Preparing terminological definitions for indexing and retrieval thesauri: a model. // *Advances in Knowledge Organization*. Würzburg: Ergon Verlag. 5, 363-369.
- International Standard Organization (2011). International Standard ISO 25.964-1: Information and Documentation: Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval. Geneva: ISO.
- Jackson, Peter; Moulinier, Isabelle (2007). Natural language processing for online applications: text retrieval, extraction and categorization. 2nd. ed. rev. Philadelphia: Benjamins.
- Jorna, K.; Davies, S. (2001). Multilingual thesauri for the modern world: no ideal solution?. // *Journal of Documentation*. 57:2, 284-295.
- Marcondes, Carlos Henrique (2011). O papel das relações semânticas na organização e representação do conhecimento em ambientes digitais. // Da Silva, Fabiano Couto Corrêa; De Sales, Rodrigo, organizadores. *Cenários da organização do conhecimento: linguagens documentárias em cena*. Brasil: Thesaurus. 129-168.
- Mesquita, Luiz Antônio Lopes; Souza, Renato Rocha Souza; Porto, Renata Maria Abrantes Baracho (2014). Noun Phrases in Automatic Indexing: a Structural Analysis of the Distribution of Relevant Terms in Doctoral Theses. // 13ª Conferência Internacional (ISKO 2014, Cracovia, Polonia). 8 p. [http://www.eci.ufmg.br/documentos-arquivos/MESQUITASOUZAPORTO\\_NOUN\\_PHRASES\\_IN\\_AUTOMATIC\\_INDEXINGenglish.pdf](http://www.eci.ufmg.br/documentos-arquivos/MESQUITASOUZAPORTO_NOUN_PHRASES_IN_AUTOMATIC_INDEXINGenglish.pdf) (30-05-0214).
- Moreiro, J. A. (1994). Documentación y lingüística: conceptos de relación esenciales. // *Ciencias de la Información*. 25:4, 202-211.
- Mote, L. J. (1962). Reasons for the variations in the information needs of scientists. // *Journal of Documentation*. 18:4, 169-175.
- National Information Standards Organization (2010). Guidelines for the construction, format and management of monolingual controlled vocabularies: an American National Standard developed by the National Information Standards Organization. Bethesda: NISO Press. [http://www.niso.org/apps/group\\_public/download.php/12591/z39-19-2005r2010.pdf](http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf) (30-05-0214).
- Naumis, Catalina (2007). Los tesauros documentales y su aplicación en la información impresa, digital y multimedia. México: UNAM; Buenos Aires: Alfagrama.
- Noruzi, A. (2006). Folksonomies: (Un)controlled vocabulary?. // *Knowledge Organization*, 33(4), 199-203.
- Ogden, Charles Kay (1930). *Basic English: A General Introduction with Rules and Grammar*. London: Paul Treber.
- Palmer, Harold (1936a). The history and present state of the movement towards vocabulary control: Part 1. // *Bulletin of the Institute for Research in English Teaching*, 120, 14-17 (January 1936).
- Palmer, Harold (1936b). The history and present state of the movement towards vocabulary control: Part 2. // *Bulletin of the Institute for Research in English Teaching*, 121, 19-23 (February 1936).
- Palmer, Harold (1938). *A grammar of english words*. London: Longmans.
- Peterson, E. (2006). Beneath the Metadata: some philosophical problems with Folksonomy. // *D-Lib Magazine*. 12:11. <<http://www.dlib.org/dlib/november06/peterson/11peterson.html>> (30-05-0214).
- Rao, Manikya (1998). Natural language versus controlled vocabulary in information retrieval: A case study in soil mechanics. // *Journal of the American Society for Information Science*. 49:10, 881-887.
- Simón, Lucía et al (2013). Estructuras sintagmáticas en tesauros: aproximación cualitativo-cuantitativa. // *Actas de las 3ª Jornadas de Intercambios y Reflexiones acerca de la Investigación en Bibliotecología*, La Plata, Argentina, 28-29 de noviembre de 2013. La Plata: Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata, 2013. 1-15. [http://www.jornadabibliotecologia.fahce.unlp.edu.ar/jornadas-2013/actas-2013/simon\\_fontans\\_barite\\_duarte\\_.pdf](http://www.jornadabibliotecologia.fahce.unlp.edu.ar/jornadas-2013/actas-2013/simon_fontans_barite_duarte_.pdf) (30-05-2014).
- Slavic, Aida (2007). On the nature and typology of documentary classifications and their use in a networked environment. // *El profesional de la información*. 16:6 (nov.-dic. 2007) 580-589.
- Taube, Mortimer (1955). *Uniterm System of Indexing Operating Manual*. Washington: Documentation Inc.
- Unesco. Programa General de Información y UNISIST (1984). *Directrices para el establecimiento y desarrollo de tesauros monolingües*. 2a. ed. rev. París: Unesco.
- Vickery, B.C. (1997). Ontologies. // *Journal of Information Science*. 23:4, 277-286.
- West, M. (1953). *A General Service List of English Words*. London: Longman, Green and Co.

Enviado: 2014-06-23. Segunda versión: 2014-06-25.  
Aceptado: 2014-06-26.