

# DEFINICION: NUEVAS TENDENCIAS EN TESAURUS Y OTROS VOCABULARIOS CONTROLADOS

Amelia Aguado<sup>1</sup>  
Rosa Z. Pisarello<sup>2</sup>  
Ana M. Martínez<sup>3</sup>  
Cecilia Delorenci<sup>4</sup>

Cátedras de <sup>1</sup>Bibliografía y Selección, <sup>2</sup>Referencia General, <sup>3</sup>Clasificación, Departamento de Bibliotecología y <sup>4</sup>Cátedra de Lingüística, Departamento de Letras. Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata. Calle 48 e/6 y 7-5º piso, 1900 La Plata, Argentina. E-mail: dhubi@huma.fahce.unlp.edu.ar.

## RESUMEN

La nota de alcance en los tesauros es un recurso para desambiguar aquellos descriptores que no son suficientemente claros. Sin embargo, la concepción tradicional está siendo revisada, distinguiéndose entre definición y nota de alcance, con el fin de cerrar la brecha entre los bancos terminológicos y los tesauros y ayudar a los usuarios finales durante la búsqueda por materia en los sistemas de información en línea. En este trabajo, se analizan definiciones tomadas de fuentes terminológicas en Bibliotecología y Documentación y se concluye que la inclusión de una definición por género próximo y diferencia específica para cada descriptor se apoya en el ordenamiento sistemático del tesoro.

**Palabras clave:** Definiciones, Notas de alcance, Tesauros, Vocabularios controlados, Terminología.

## ABSTRACT

The scope note in thesauri is a resource for disambiguating those descriptors which are not sufficiently clear. However, the traditional conception is being reviewed, distinguishing definition from scope note, in order to close the gap between terminological banks and thesauri, and to help end-users during the subject search in on-line information systems. In this work the definitions from terminological sources in the field of Library Science and Documentation are analysed, and we conclude that the inclusion of a definition by genus and differentia for each descriptor is supported by the classified or systematic arrangement of the thesaurus.

**Keywords:** Definitions, Scope notes, Thesauri, Controlled vocabularies, Terminology

## INTRODUCCION

Como es sabido, la norma ISO 2788 (1) para la construcción de tesauros monolingües y sus equivalentes nacionales, así como las directrices de IFLA para los registros de autoridad y referencia de materia (GSARE) (2) y los reconocidos manuales de construcción de tesauros de Lancaster (3) y Aitchison y Gilchrist (4) establecen un recurso para la desambiguación de aquellos descriptores cuyo significado no es suficientemente claro, mediante la inclusión de una nota de alcance. Aunque las notas de alcance pueden ofrecer definiciones limitadas o extensas, una definición completa tipo diccionario no suele incluirse en los tesauros, sino que se recomienda únicamente para los términos prestados, los neologismos o los términos que se utilizan en forma diferente a lo común.

Sin embargo, esta concepción ha comenzado a revisarse en los últimos años. La nueva edición del manual de Aitchison et al. (5) distingue claramente entre definición y nota de alcance y señala una preocupación actual por incrementar el número de definiciones en los tesauros. ¿Por qué este cambio? Los mismos autores ofrecen como una posible explicación la tendencia a cerrar la brecha entre los bancos terminológicos y los tesauros, que exige la inclusión de una definición para cada descriptor. Sin embargo, se puede afirmar que la necesidad de incluir una definición, desde el punto de vista de la recuperación de información, se relaciona también con

las dificultades que se han detectado en las búsquedas por materia en los sistemas de información en línea (6-7).

A pesar de las normas, directrices y manuales ya citados (1-4), los actuales tesauros no siempre se caracterizan por su buena calidad y a veces se hace difícil identificar correctamente el descriptor necesario para la indización o la búsqueda, especialmente en el área de las ciencias sociales y las humanidades. Raya Fidel (8), de la University of Washington at Seattle, realizó una investigación en 15 bases de datos de distintas disciplinas, con la finalidad de conocer las preferencias de los usuarios respecto al uso de un tesoro o de la lengua natural para realizar una búsqueda en línea, obteniendo los siguientes resultados: a) 25% de los usuarios ni siquiera intentaron seleccionar un descriptor, manifestando que la búsqueda por lengua natural era la forma más eficiente de buscar información, b) 25% de los usuarios buscaron descriptores en el tesoro, pero no encontraron ninguno que coincidiera con su tema de búsqueda y c) 50% de los usuarios localizaron un descriptor que les permitiera recuperar información. Podría plantearse entonces que la posibilidad de contar con una definición para cada descriptor permitiría tener términos claros y unívocos a disposición tanto de indizadores como de usuarios.

Se han señalado dos opciones para las definiciones en los tesauros. La primera consiste en conectar el tesoro con uno o varios diccionarios, de manera que cada descriptor esté vinculado a una definición, ya sea por algún tipo de nexo en los tesauros impresos, o por un enlace entre versiones informáticas; ejemplos de este tipo son The computer database thesaurus and dictionary, The management contents database thesaurus and dictionary y NASA thesaurus (9). La dificultad más importante de estos sistemas es vincular cada descriptor, diferenciado de sus posibles homónimos y polisemas, con una única acepción.

La segunda opción es incluir una definición en el tesoro para cada uno de los descriptores, como ocurre en el Energy database subject thesaurus (9). Mientras que Svenonius (10) recomienda la inclusión de definiciones lo más amplias posibles, otros autores prefieren recomendar una forma acorde con la estructura del tesoro, mediante una definición por género próximo y diferencia específica (11-13). Desde 1994, Sager y L'Homme (11) Sager y Ndi-Kimbi (12) y Michèle Hudon (13) han propuesto distintos modelos para construir la definición de los descriptores, basados de distintas facetas.

Según la norma ISO 1087, la definición es un "enunciado que describe un concepto y permite su diferenciación de otros conceptos dentro de un sistema de conceptos". Al incluir el término enunciado, esta formulación implica un enfoque terminológico del problema: Precisamente, la terminología apunta a la representación de los conceptos, es decir la mejor manera de identificar un concepto dado por medio de palabras o frases. Por otra parte, el terminólogo se interesa también por el modo en que los conceptos se vinculan en forma sistemática, por lo cual prestan atención a los sistemas de clasificación, sobre todo porque en muchos casos la única forma de definir un concepto consiste en ubicarlo dentro de un contexto. Sin embargo, aunque el resultado parezca idéntico, la tarea del terminólogo se diferencia de la labor del documentalista: El terminólogo apunta a definir términos usados en una disciplina determinada. A partir de los textos especializados correspondientes, establece relaciones entre esos términos y a posteriori llega a una suerte de clasificación, en la que un término designa a una especie incluida en un término más general o a una parte de un todo, o bien se caracteriza por una función. En cambio, el que construye un tesoro de indización necesita vincular una parte alfabética y una parte sistemática. La parte alfabética desmembra a la sistemática y explicita qué descriptores son más amplios, cuáles están relacionados entre sí, cuáles se excluyen. La parte sistemática expresa relaciones jerárquicas entre conceptos y elige designaciones unívocas para esos conceptos. En la construcción de tesauros, por lo tanto, la aclaración del concepto es previa a la determinación del término elegido o creado para designarlo, así como es previa a su inclusión en el sistema clasificatorio que luego se disgrega en el ordenamiento alfabético de los descriptores de cada concepto.

Según Riggs (14), para identificar un concepto dentro de este sistema se requiere describir cada concepto por separado y mostrar cómo se relaciona con otros conceptos pertenecientes al mismo sistema. El autor señala que la onomasiología, disciplina que a menudo se identifica con uno de sus aspectos, la onomástica (estudio de los nombres propios personales y de los topónimos), no se ha dedicado suficiente atención al subcampo que

entiende en la designación de los conceptos. Propone el término onomántica para denominar ese subcampo. El resultado del trabajo onomántico se traduce en un sistema en que cada concepto adquiere una denominación unívoca, para la que se conviene una definición. Vale decir que, sin que haya una prescindencia total del sistema lingüístico, puesto que los términos clasificatorios proceden de él o se construyen de conformidad con sus normas morfológicas y sus raíces semánticas, la organización del conocimiento es el fundamento de las relaciones expresadas mediante términos. El sistema logrado, por lo tanto, posee una fuerte dosis de artificialidad. Para constituir un sistema de esta índole se requiere una precisa definición de cada uno de los términos utilizados. El trabajo de Mönke (15) enumera más de 70 tipos de definiciones, si bien los más importantes para la asignación de términos en los tesauros se reducen fundamentalmente a dos: por comprensión y por extensión.

En la definición por comprensión se parte de un concepto supraordinado (expresado por un hiperónimo) y se indican las características específicas que distinguen al concepto por definir de otros de la misma serie (la expresión de ese concepto es un hipónimo). En ese sistema de definiciones, que se basa en relaciones lógicas, no es necesario enumerar todas las características conocidas de un concepto: basta con las consideradas esenciales según la finalidad de la definición. La estructura del sistema, en que cada concepto se relaciona con el supraordinado correspondiente, resulta en un encadenamiento de conceptos que va del más general al más particular. Hiperónimos e hipónimos se encadenan entre sí, de modo que un término general en una definición puede ser término específico en otra. (Por ejemplo, animal es un ser vivo dotado de movimiento, mientras que hombre es un animal racional; en el primer caso, se incluye a animal dentro de la categoría más amplia de los seres vivos, mientras que en el segundo pasa a ser la categoría dentro de la cual está incluido hombre).

Sin duda, la definición por comprensión es la más apta para los tesauros, dado que los sistemas conceptuales generados a partir de ella se corresponden con la estructura jerárquica de su ordenamiento sistemático. Aún así, conviene caracterizar otros tipos de definición susceptibles de aplicación en la construcción de tesauros. En la definición por extensión, pueden enumerarse todos los conceptos subordinados que se encuentran en el mismo nivel de clasificación (que se expresarían en un tesoro como términos fraternos o intracategorías) o bien todos los objetos designados con el término (aunque este último caso sólo es válido cuando la definición abarca un número limitado de objetos, por ejemplo los planetas del sistema solar). En otros casos, se puede indicar la regla que permite llegar a la enumeración (por ejemplo, en casos como número par es el divisible por dos, pero esta modalidad definitoria puede asimilarse con facilidad a la definición por comprensión). La clásica introducción de Arntz y Picht (16) señala además definiciones genéticas y operacionales (en ellas se relacionan las diferentes operaciones necesarias para determinar el definiendum; si las operaciones se consideran características, este tipo de procedimiento puede asimilarse a la definición por comprensión, mientras que si se las equipara a conceptos subordinados, se parecen a una definición por extensión), definiciones por el contexto (de escasa utilidad, salvo para aclarar conceptos) y definiciones nominales (proporcionan sinónimos de los términos; en un tesoro, equivaldrían a los términos no usados).

En el presente trabajo analizamos las definiciones utilizadas en diversas fuentes terminológicas, con el propósito de evaluar el uso de la definición por género próximo y diferencia específica para los descriptores de un tesoro.

## **METODOLOGIA**

Se seleccionó una muestra de términos de distintas fuentes terminológicas en Bibliotecología y Documentación, bajo las siguientes consideraciones, tomando como parámetro de referencia la norma ISO 5127/3a (17) sobre documentos tradicionales. Como complemento, se utilizó también la norma española UNE 50113/3a (18), que constituye prácticamente una traducción al español de la norma ISO ya mencionada. Es importante señalar que ambas normas incluyen los términos y sus definiciones en un ordenamiento sistemático, de acuerdo con un esquema de clasificación propio. También se analizaron los glosarios de Wersig y Neveling (19), Harrod (20), ALA (21) y Buonocore (22), el primero de los cuales también presenta los términos y definiciones en un ordenamiento sistemático. Por último también se extrajeron definiciones del Diccionario de la lengua española (23) y del de María Moliner (24), que por supuesto poseen un ordenamiento alfabético. En cuanto a los

vocabularios controlados se estudiaron el Vocabulario controlado de bibliotecología y ciencias de la información (VCBCI) de Peniche Sánchez McGregor (25) y ASIS thesaurus of information science and librarianship (ASIS) (26). El primero es una lista de descriptores que no incluye relaciones jerárquicas, mientras que el segundo sí es un tesoro, en el que todos sus descriptores están organizados en un ordenamiento sistemático.

Una vez reunidas las fuentes, se completaron los siguientes pasos: Se extrajeron 78 términos de la norma ISO 5127/3a, así como 506 definiciones de todas las fuentes consultadas: 156 de las dos normas, 237 de los cuatro glosarios y 113 de los dos diccionarios. Se analizaron todas las definiciones para establecer cuáles de ellas utilizaban la definición por género próximo y diferencia específica. Se extrajeron los términos equivalentes de los dos vocabularios controlados seleccionados y se analizaron sus correspondientes notas de alcance y relaciones jerárquicas, para comparar su correspondencia con el tipo de definición elegido.

## RESULTADOS Y DISCUSION

De las 156 definiciones extraídas de las dos normas, 121 (77%) fueron por género próximo y diferencia específica, mientras que 93 de las 237 definiciones extraídas de los cuatro glosarios (39%) y 33 de las 113 obtenidas de los diccionarios (29%) utilizaban el mismo tipo de definición. Resulta evidente, entonces, que las normas muestran una preferencia mucho mayor por la definición de género próximo y diferencia específica en comparación con los glosarios y diccionarios; sin embargo, la proporción que presentan estos dos últimos (aproximadamente un tercio en ambos casos) es suficientemente importante para confirmar la validez y el uso de este tipo de definiciones. Es importante señalar que el 51% de las definiciones incluidas en Wersig y Neveling, 38% de las de Harrod, 49% de las de ALA y 32% de las de Buonocore fueron por género próximo y diferencia específica. De manera que, si se tiene en cuenta que tanto las normas como el glosario de Wersig y Neveling poseen un ordenamiento sistemático (característico también de los tesauros), nuestros resultados apoyarían la opinión expresada en la Introducción de que este tipo de ordenamiento favorecería la construcción de definiciones por género próximo y diferencia específica.

Por otra parte, del total de 78 términos extraídos de la norma ISO 5127/3a (17), sólo 26 (36%) estaban incluidos en los dos vocabularios controlados, mientras que 23 términos (29%) se encontraban solamente en uno. Otros 29 términos (37%) no se incluían en ninguno de los dos vocabularios. De los 45 términos incluidos en VCBCI, 14 (31%) incluían una nota de alcance, siendo 13 de ellas definiciones, aunque sólo 3 por género próximo y diferencia específica; como ya se explicó, ninguno de los descriptores contaba con relaciones jerárquicas. En contraste, de los 30 términos incluidos en ASIS, sólo uno incluía nota de alcance con instrucciones para el indizador y no como definición; sin embargo, todos los descriptores mostraban términos genéricos y 11 de ellos (37%) términos específicos.

Estos resultados coinciden con los de Alvaro Bermejo et al. (27) en cuanto a la escasa cantidad de descriptores que incluyen notas de alcance. El estudio realizado por este grupo de investigadores españoles, si bien no establece una diferencia clara entre definición y nota de alcance, llama la atención sobre la alta tasa de ambigüedad de distintos tesauros en lengua española, ya que de una muestra de 47 tesauros, dos tercios sólo incluían notas de alcance en menos del 10% de los descriptores, destacándose, como una excepción, el caso del tesoro de Psicología (28) con 65% de descriptores acompañados de notas de alcance.

Debe recordarse que el VCBCI no posee una estructura jerárquica, mientras que todos los descriptores de ASIS poseen una indicación de término genérico y/o término específico. La comparación de esta característica con la disponibilidad de notas de alcance, en ambos vocabularios controlados, parece sugerir que la estructura jerárquica haría innecesaria la inclusión de una definición por género próximo y diferencia específica y vice versa, la falta de una estructura jerárquica requeriría mayor cantidad de estas definiciones. Sin embargo, la forma narrativa de la definición puede facilitar la comprensión por parte de los indizadores y usuarios finales de los sistemas de información; en este sentido, Buchan (9) considera que la combinación de la estructura sistemática del tesoro con la definición puede resultar enriquecedora, afirmando que: "A partir de la estructura del tesoro se puede redactar una definición. A partir de una definición se puede comprender mejor la jerarquía de un descriptor".

## CONCLUSIONES

Los tesauros y otros vocabularios controlados, que antes eran propiedad exclusiva del indizador, se han convertido en herramientas de recuperación útiles para los usuarios de sistemas de información en línea, a quienes se debe proporcionar la mayor cantidad de información posible para facilitar la elección de los términos de búsqueda. Una contribución importante en este sentido es la inclusión de una definición para cada uno de los descriptores u otros términos de indización. La definición por género próximo y diferencia específica es ampliamente utilizada en las fuentes terminológicas y puede adaptarse sin mayores dificultades a los tesauros, debido a que el ordenamiento sistemático de éstos parece facilitar su redacción.

No obstante, como se mencionó anteriormente se requiere continuar con este tipo de investigaciones para establecer la utilidad de otros tipos de definición que contribuyan a enriquecer la comunicación entre un sistema de información y sus usuarios, eliminando la ambigüedad y la vaguedad de los términos, así como el posible costo-beneficio de mantener un tesoro con información tan detallada.

## AGRADECIMIENTOS

Esta investigación ha sido subsidiada por el Programa de Incentivos del Ministerio de Educación de la República Argentina.

## BIBLIOGRAFIA

- [1] ISO 2788. "Documentation: Guidelines for the establishment and development of monolingual thesauri". 2 ed. Geneva: ISO, 1986.
- [2] IFLA. "Guidelines for subject authorities and reference entries". München: KG Saur, 1993.
- [3] Lancaster, F.W. "Vocabulary control and information retrieval". 2 ed. Arlington, Va: Information Resources Press, 1986.
- [4] Aitchison, J. and Gilchrist, A. "Thesaurus construction". 2 ed. London: Aslib, 1987.
- [5] Aitchison, J., Gilchrist, A. and Bawden, D. "Thesaurus construction". 3 ed. London: Aslib, 1997.
- [6] Holley, R.L. "Subject access in the online catalog". New York: Haworth Press, 1989.
- [7] Larson, R.R. The decline of subject searching: Long term trends and patterns of index use in an online catalog. "Journal of the American Society for Information Science", 42(4), 1991, 197-215.
- [8] Fidel, R. Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. "Journal of the American Society for Information Science", 42(7), 1991, 401-414.
- [9] Buchan, R.L. Intertwining thesauri and dictionaries. "Information services and use", 9, 1989, 171-175.
- [10] Svenonius, E. Definitional approaches in the design of classification and thesauri and their implications for retrieval and for automatic classification. En: "Knowledge organization for information retrieval: Proceedings of the 6th International Study Conference on Classification Research", University College, London, 16-18 June 1997.
- [11] Sager, J.C. and L'Homme, M.C. A model for the definition of concepts rules for analytical definitions in terminological databases. "Terminology", 1, 1994, 351-353.
- [12] Sager, J.C. and Ndi-Kimbi, A. The conceptual structure of terminological definitions and their realisations: A report on research in progress. "Terminology" 2, 1995, 61-81.
- [13] Hudon, M. Preparing terminological definitions for indexing and retrieval thesauri: a model. En: Knowledge organization and change: Proceedings of the 4th International ISKO Conference, Washington DC, 15-18 July 1996. "Advances in knowledge organization", 5, 1996, 363-369.
- [14] Riggs, F.W. Onomastics and terminology. Pt. I: Their contributions to knowledge organization. "Knowledge Organization"; 23(1), 1996, 25-33.
- [15] Mönke H. Definitionstypen und Definitionsmatrix. "Nachrichten für Dokumentation", 29(2), 1978, 51-60.
- [16] Arntz, R. and Picht, H. "Introducción a la terminología". Madrid: Fundación Germán Sánchez Ruipérez, 1995.

- [17] ISO 5127/3<sup>a</sup>. "Documentation and information: Vocabulary: Traditional documents". Geneva: ISO, 1983.
- [18] UNE 50113/3a. Documentación e información: Vocabulario: Documentos tradicionales. Madrid: AENOR, 1983.
- [19] Wersig, G. and Neveling, U. "Terminology of documentation". 2 ed. París: Unesco, 1976.
- [20] Harrod, L.M. "The librarian's glossary of terms used in librarianship, documentation and the book crafts and reference book". 5 ed. New York: Bowker, 1990.
- [21] Young, H. "Glosario ALA de bibliotecología y ciencias de la información". Madrid: Díaz de Santos, 1983.
- [22] Buonocore, D. "Diccionario de bibliotecología: términos relacionados a la bibliología, bibliografía, bibliofilia, biblioteconomía, archivología, documentología, tipografía y materias afines". 2 ed. Buenos Aires: Marymar, 1980.
- [23] Moliner, M. "Diccionario de uso del español". Edición en CDROM. Madrid: Gredos, 1996.
- [24] Real Academia Española. "Diccionario de la lengua española". 21 ed. Madrid: Espasa Calpe, 1992.
- [25] Peniche Sánchez McGregor, S. "Vocabulario controlado en bibliotecología y ciencia de la información". México: CUIB, UNAM, 1992.
- [26] Milstead, J.L. "ASIS Thesaurus of information science and librarianship". Silver Spring: ASIS, 1994.
- [27] Alvaro Bermejo, C., Villagrà Rubio, A. y Sorli Rojo, A. Desarrollo de lenguajes documentales formalizados en lengua española. II. Evaluación de los tesauros disponibles en lengua española. "Revista española de documentación científica", 12(3), 1989, 283-297.
- [28] García, M.B. y Monfasani, R.E. "Construcción de un tesoro de psicología". Buenos Aires: UBA, 1983.