

Las relaciones semánticas: su papel en la materialización de la representación y organización del conocimiento

M^a José López-Huertas Pérez
Universidad de Granada

0.1. Resumen

La semántica de los lenguajes documentales es uno de los aspectos más interesantes y menos evolucionados de los mismos. Se desarrolla una metodología nueva, dando ejemplos concretos, para identificar relaciones semánticas pertinentes y exhaustivas para un dominio temático determinado a partir de los textos especializados en ese tema.

Palabras clave: Relaciones semánticas. Lenguajes documentales. Metodologías.

0.2. Abstract

Semantics of documentary languages is one of the most important issues of this matter, nevertheless it is not well developed. It needs more attention in order to answer actual information retrieval systems demands. A new methodology is developed, giving actual data, that allows to identify pertinent and exhaustive semantic relations taken from specialized texts for a giving subject domain.

Keywords: Semantic relations. Indexing languages. Methodologies.

1. Introducción

De todos los aspectos relacionados con los lenguajes documentales, tal vez sea el semántico uno de los más importantes. Sin duda, por ello ha sido objeto de estudio continuo desde los años sesenta hasta nuestros días. Gardin incide especialmente en la importancia de este aspecto y comenta la necesidad de estudiar las modernas teorías del significado a fin de hacer más eficaces los sistemas de recuperación de la información (Gardin, 1979). Asimismo, la semántica está en estrecha relación con el concepto de relevancia y éste con el de evaluación, razones por las que la semántica es un tema crucial para los sistemas de recuperación de la información. Así se reconoce en Robertson (Robertson, 1979), Saracevic

(Saracevic, 1975) y más recientemente en Ingwersen (Ingwersen, 1992) y Schamber (Schamber, 1994), entre otros.

2. El aspecto semántico de los lenguajes documentales

El interés que suscita el aspecto semántico de los lenguajes documentales está en estrecha relación con la importancia que se le concede a la semántica de los sistemas de información en general, como se acaba de decir, porque los lenguajes documentales son una de las principales herramientas de las que se valen los sistemas para cumplir sus funciones de comunicación y, por tanto, en la medida en que la semántica documental cumpla adecuadamente sus funciones de representación y organización del conocimiento a través de las estructuras conceptuales de los lenguajes, estos incrementarán las capacidades de comunicación de los sistemas a los que sirven, haciéndolos así más eficaces. En la medida en que los lenguajes documentales sean capaces de representar y organizar adecuadamente el conocimiento desempeñarán mejor sus funciones en los sistemas de información. Como quiera que la representación y organización del conocimiento es, a mi modo de ver, uno de los principales papeles que tienen asignado los lenguajes y lo que les permite ser mediadores e instrumentos de la comunicación documental, cualquier tema relacionado con estas funciones de los lenguajes documentales suscita un gran interés.

En los lenguajes documentales, la representación del conocimiento se produce a dos niveles. Por un lado, distinguimos la expresión verbal de la representación cuyo principal agente es el descriptor, considerado tanto desde el punto de vista de su forma de expresión gráfica (notaciones, descriptores naturales) como de su significado (semántica intralingüística, al nivel del descriptor). Por otro lado, tenemos la expresión conceptual de la representación (estructura conceptual propiamente dicha) que es la semántica al nivel del lenguaje, que a su vez constituye una propuesta de organización del conocimiento. Aunque ambos niveles son necesarios, este trabajo se centrará básicamente en el segundo, el estructural, porque es el directamente relacionado con el tema que aquí se estudia: las relaciones semánticas y su papel en la representación y organización del conocimiento.

Si partimos de la idea de que el conocimiento público está depositado en los documentos y las estructuras conceptuales de los lenguajes documentales deben ser, en parte, una réplica del contenido de esos documentos, es decir la semántica de los mismos (López-Huertas, 1997), no cabe duda de que la semántica de los lenguajes documentales tiene asignada la función de representar el conocimiento público a través de la semántica de los documentos. Por tanto, el proceso de la representación se origina, en parte, en la semántica de los documentos especializados —su significado—. Y digo en parte porque no puede olvidarse que el len-

guaje bien diseñado también debe representar la semántica —el significado— de las preguntas de los usuarios al sistema de información. Los lenguajes documentales, por tanto, representarán ambas semánticas, la de los documentos y la de los usuarios. En el primer caso el lenguaje representará el conocimiento depositado por los generadores de la información en los documentos y en el segundo caso será el conocimiento que le falta al usuario para realizar una determinada tarea. Asimismo, una vez elaborado el lenguaje, éste se utiliza para representar los documentos del sistema (su semántica) y las preguntas de usuarios (su semántica) por medio de una operación llamada indización.

Por todo lo dicho, la representación y organización del conocimiento por medio de los lenguajes documentales se hace básicamente a través de sus estructuras conceptuales que se fundamentan en unas bases eminentemente semánticas. Esta representación se materializa y se hace funcional por medio de las relaciones semánticas que se establecen en el mismo lenguaje documental que permiten relacionar todo su vocabulario, posibilitando la utilización por parte de sus usuarios de la organización del conocimiento que el lenguaje propone. Las relaciones semánticas van a permitir que los usuarios puedan conocer y utilizar la representación y organización del conocimiento que propone el sistema de información que están utilizando para resolver su falta de conocimiento sobre una materia dada. Establecen, o por lo menos así debería ser, una especie de redes neuronales, cuanto más densas mejor, que hacen posible que los usuarios puedan navegar por toda la estructura, aprovechando el símil biológico otra vez, de forma similar a como los impulsos eléctricos navegan por el cerebro. Por tanto, la importancia de esos elementos conductores que son las relaciones semánticas o relaciones conceptuales es verdaderamente notable.

Probablemente, lo que acabo de decir no se parece mucho a la realidad de los lenguajes documentales al uso que todos conocemos que se reduce a una escasísima y poco útil enumeración de vagas relaciones conceptuales. Precisamente este hecho es lo que hace que esté hablando ahora de este tema, porque ha llegado la hora de preguntarse qué es lo que usamos y qué es lo que diseñamos y creamos cuando hacemos un lenguaje documental nuevo, sobre todo un tesoro que es el lenguaje que se elabora más frecuentemente en la actualidad. Lo que a partir de ahora voy a exponer va a tener como principal objetivo rellenar la laguna conceptual existente entre las relaciones semánticas establecidas de los lenguajes documentales tal y como ahora las conocemos y las relaciones semánticas vistas desde la perspectiva antes apuntada. Este recorrido va a conducir ineludiblemente a una realidad y a un concepto distintos de lo que normalmente se entiende por relaciones semánticas.

3. Necesidad de una nueva metodología

Para empezar es necesario volver a los orígenes, a donde todo comienza, que como se ha dicho antes son los documentos como depositarios del conocimiento público. Si, según ha quedado establecido, las estructuras conceptuales de los lenguajes documentales deben representar y organizar el conocimiento a través de la semántica de los textos, las relaciones semánticas que unen todo ese entramado conceptual deben también originarse o basarse en los textos fundamentalmente. De esta forma, el equipo encargado de hacer el lenguaje documental no tendrá que “inventarse” las clases de relaciones a establecer y todo el conjunto del lenguaje no será otra cosa que una proyección de lo que los textos consideren pertinente en relación con una materia determinada con lo que quedarán automáticamente validadas. Las repercusiones de esta manera de hacer son importantes y hablaremos de ellas más adelante.

Para poder identificar relaciones semánticas a partir de los textos, es conveniente partir de documentos especializados, que en este caso van a ser sobre instrumentos musicales. También hay que tener en cuenta que cuando estamos en este momento de elaborar un lenguaje, ya disponemos de un *stock* de vocabulario refinado que representa el tema que estamos tratando. A partir de aquí, lo que buscamos es saber lo que significa cada uno de los conceptos representados por nuestro vocabulario para los especialistas en la materia para la que se hace el lenguaje. Uno de los fundamentos del método reside en la creencia de que cuando un especialista en una determinada materia, por ejemplo, habla o describe un concepto está dándonos una gran cantidad de información relevante sobre el mismo que puede ser utilizada no sólo para la identificación y organización de categorías en la estructura (López-Huertas, 1997, 1998) sino también para la identificación de los caminos conceptuales que conectan la estructura, que es lo que conocemos como relaciones semánticas.

Si, por ejemplo, en las definiciones de los instrumentos encontramos alusiones a la procedencia del sonido (instrumento de cuerda), a la forma de producir el sonido (las cuerdas se puntean), al medio utilizado para producir el sonido (las cuerdas se puntean con la púa), al carácter musical (instrumento de profunda sonoridad), a sus características físico-morfológicas (es triangular, circular, tiene cuello, trastes, etc.), al material del que está elaborado (madera, piel, etc.), a los géneros musicales que se tocan con él (se toca música de cámara, etc. con él), a las formas musicales (se utiliza para tocar sonatas), los compositores que han escrito partitura para él (Beethoven escribió partituras para violonchelo), los intérpretes, métodos de aprendizaje, los lugares de procedencia y difusión, el tiempo en el que sitúa el instrumento musical, los grupos étnicos que lo utilizan, los acontecimientos sociales o estacionales con los que están relacionados, etc., nos están poniendo sobre la pista de las relaciones conceptuales potenciales para

el grupo temático “instrumentos musicales”. Una muestra, aunque incompleta, de la materialización de estas relaciones sería la siguiente:

- Procedencia del sonido.
- Cómo se produce el sonido.
- Qué medios se utilizan para producir el sonido.
- Carácter musical.
- Géneros musicales relacionados con él.
- Formas musicales relacionadas con él.
- Conjuntos musicales relacionados con él.
- Orquestas relacionadas con él.
- Relación con otros instrumentos.
- Cómo es físicamente.
- De qué material está elaborado.
- Con qué grupos étnicos, militares o religiosos se relaciona.
- Acontecimientos sociales/estacionales/litúrgicos.
- Lugar del que procede o relaciona.
- Tiempo.
- Compositores.
- Intérpretes.
- Métodos de aprendizaje.

Pues bien, todas estas son las relaciones semánticas que pueden adjuntarse a cada uno de los instrumentos musicales. Si tenemos en mente cómo es un tesoro, por ejemplo, los descriptores serían los nombres de los instrumentos musicales y las relaciones semánticas serían, entre otras, las que se acaban de enumerar más arriba. El cambio que se observa es casi notable porque en la propuesta que aquí se hace se producen tres modificaciones básicas con respecto a lo que viene haciéndose:

- *Primera*: no existen designaciones genéricas del tipo TR, TG o TN que dan entrada a un número de descriptores de ambigua relación conceptual, sobre todo en el caso de las relaciones asociativas. En mi anterior propuesta, muy al contrario, se enumeran relaciones conceptuales con nombre propio, identificadas y, por tanto, ausentes de ambigüedad, auténticamente controladas y pertinentes.
- *Segunda*: el número de relaciones semánticas que se establecen, siendo muchísimo más elevado, en el caso que estamos comentando, que en los

lenguajes al uso, donde sólo hay dos clases: jerárquicas y asociativas porque las de equivalencia no se generan al nivel de estructura sino de vocabulario. Si tenemos en cuenta que la necesidad de identificar mayor número de relaciones conceptuales en los lenguajes documentales es algo reiteradamente manifestado y demandado por los especialistas, los profesionales y los usuarios, este nuevo enfoque teórico y metodológico permite dar respuesta a esta demanda, al identificar bastantes más clases de relaciones individualizadas. Al mismo tiempo, al elevar considerablemente el número de relaciones semánticas, se está introduciendo en el lenguaje un potente elemento de exhaustividad controlada con el consiguiente beneficio para la recuperación de la información. A continuación se muestran las relaciones previstas para el descriptor *guitarra* en los *Encabezamientos de materia de la OEA* (Encabezamientos de la OEA, 1985), en los *Encabezamientos de la Biblioteca del Congreso* (LCSH, 1991) y las que se proponen en este trabajo, donde se observa claramente las diferencias señaladas (Fig. 1).

La riqueza semántica que se deriva del método señalado sirve también para otros interesantes propósitos como son el diseño y la recuperación de la información en sistemas automatizados y, principalmente, el diseño de sus interfaces. El sistema puede utilizar las rutas conceptuales marcadas por las relaciones semánticas, del tipo de las antes descritas, para conducir al usuario por un viaje guiado con precisión y pertinencia a través de la estructura del lenguaje. Podría ser una forma eficaz de asistir al usuario en sus búsquedas y una alternativa a los escasos recursos conceptuales con que cuentan los actuales sistemas de información. Ayudarían, en definitiva, a resolver una parte importante de los problemas que tienen planteados actualmente estos sistemas.

- *Tercera*: las fuentes y el método utilizados para identificar las relaciones semánticas son básicamente distintos a los habituales. En los tesauros al uso las relaciones semánticas se suelen trazar de acuerdo con lo que sugiere el vocabulario y de acuerdo con una nutrida lista de relaciones conceptuales, que existen ya establecidas, para el caso de lo que se vienen llamando relaciones asociativas. En cualquier caso, quien hace el tesauro decide las que son y cómo son. Muy al contrario, en el ejemplo que comentamos en este trabajo, la fuente son los textos o documentos especializados, que son usados directamente para identificar los nexos conceptuales. La intervención del equipo o la persona que elabora el lenguaje se ciñe a trasladar al mismo lo que indican los textos, reduciéndose de esta manera interpretaciones, y cualquier otro elemento de subjetividad no deseada en dicho lenguaje. Las consecuencias de esta forma de proceder son importantes.

PROPUESTA	OEA	LCSH
VIOLIN		
procedencia del sonido cómo se produce el sonido qué medios se utilizan para producir el sonido carácter musical géneros musicales relacionados con él violín tenor formas musicales relacionadas con él conjuntos musicales relacionados con él orquestas relacionadas con él relación con otros instrumentos cómo es físicamente de qué material está elaborado con qué grupos étnicos, militares o religiosos se relaciona acontecimientos sociales, estacionales y litúrgicos lugar del que procede o relaciona tiempo compositores intérpretes métodos de aprendizaje	xx instrumentos de cuerda instrumentos musicales	UF fidula TG instrumentos de cuerda frotados TN pochette

Fig.1. Relaciones previstas para el descriptor "guitarra"

Por un lado, las relaciones identificadas directamente a partir de los textos especializados pueden considerarse automáticamente validadas —no hay que demostrar su validez o idoneidad para el ámbito temático del que se trate porque los mismos textos las proclaman—. Por otro lado, y como consecuencia de lo anterior, el uso de las fuentes textuales permiten la utilización de un método de trabajo que previene en gran medida la introducción de posibles distorsiones —falsas, innecesarias o poco pertinentes asociaciones conceptuales— en la identificación e interpretación de esas relaciones, favorecidas por la intervención de quien elabora el lenguaje en su papel de ser él o la que decida sobre este asunto, cosa que es probable con el método tradicional. En definitiva, que si la estructura conceptual de los lenguajes documentales debe ser en buena parte una proyección de los textos, evitando lo más posible su manipulación por parte de los que elaboran estos lenguajes, las relaciones semánticas deben asimismo extraerse de los textos en parte, como antes se ha dicho, para que cumplan su verdadera función que es la de ser auténticas rutas conceptuales que permiten al usuario del lenguaje la materialización de la representación y organización del conocimiento propuesta por la estructura del lenguaje documental cada vez que las usa para recuperar información.

4. Conclusiones

Parte de lo hasta ahora dicho está en sintonía con una de las últimas propuestas paradigmáticas para el campo de la Biblioteconomía y Documentación, conocida como la teoría del análisis del dominio temático (Horland y Albrechtsen, 1995), que defiende, *grosso modo*, que el dominio temático debe ser el punto de referencia para el estudio de los problemas que tiene planteados esta disciplina. Desde el principio de este trabajo, se ha insistido en la necesidad de partir del dominio temático especializado para elaborar tanto la estructura conceptual como las relaciones semánticas de un lenguaje documental, porque cada tema va a proyectar una estructura conceptual propia y exclusiva y, por tanto, va a tener también unas relaciones semánticas propias y exclusivas al tiempo que diferentes de otros dominios temáticos. Por eso creo poco útil la enumeración *universal* apriorística de relaciones conceptuales, del tipo: relación entre los objetos y sus propiedades, relación entre las cosas y sus disciplinas objeto de estudio, etc. No es que no se pueda o se deba hacer una lista de posibles relaciones conceptuales asociativas, de hecho es la forma habitual de proceder. Lo que digo es que por sí solas son poco útiles para la recuperación de la información porque no representan con precisión las necesidades de nexos conceptuales de una determinada disciplina, sino que, por el contrario, están pensadas para que sean válidas para cualquiera de ellas. Tampoco quiero decir que estas relaciones

universales sean inútiles o que haya que despreciarlas. Sencillamente, podrían ser utilizadas de forma complementaria, si se considera oportuno, junto con las propias de cada campo que serían extraídas en su mayor parte de los textos especializados, según antes se ha dicho. Así considerado, efectivamente, puede hablarse de que el dominio temático, representado en los textos especializados, es el referente sobre el que se construyen las estructuras conceptuales y sus relaciones semánticas y en este punto es donde realmente veo que esta propuesta camina en una dirección parecida a la que sugiere Hjørland, convirtiéndose en un punto de apoyo para ese paradigma.

Por último, es necesario preguntarse por los beneficios que las relaciones semánticas, dentro del contexto conceptual y metodológico expuesto en este trabajo, puede reportar a los sistemas de información y, concretamente a la recuperación de la información. Porque no sólo hay que investigar alternativas dirigidas a aumentar el número de relaciones conceptuales que pueden establecerse en una determinada estructura, sino que también es necesario qué repercusión previsible puede tener este hecho en el contexto para el que han sido creadas. Las relaciones semánticas son un elemento de exhaustividad en la recuperación de la información, reconocido al menos desde Gilchrist (Gilchrist, 1971). En la medida en que se eleve el número de estas relaciones, aumentaremos la exhaustividad de los sistemas de recuperación de la información. Si según ha quedado demostrado, este número se incrementa considerablemente, estaremos con ello contribuyendo a una mejora del sistema. Este aspecto es realmente importante porque tanto los sistemas de información convencionales (Lancaster, 1991) como los no convencionales (Blair y Maaron, 1985; Blair, 1986; Maaron, 1988) están teniendo un nivel de exhaustividad muy bajo con la consiguiente pérdida de información que ello supone. Siendo importante esta faceta, no lo es menos el hecho de que las relaciones semánticas del tipo de las identificadas en este trabajo tendrían, al mismo tiempo, la cualidad de ser altamente pertinentes porque identifican e individualizan los tipos de relaciones que son relevantes para un determinado dominio temático y porque proceden directamente de los textos especializados. En consecuencia, se puede decir que las relaciones semánticas pueden ser al mismo tiempo exhaustivas y pertinentes en una medida muy semejante. Este hecho está en relación directa con las últimas tendencias en el estudio de estos dos parámetros que dicen que exhaustividad y precisión no tienen necesariamente que ser inversamente proporcionales porque ha sido demostrado que pueden aumentar o disminuir al mismo tiempo, según se aprecia en los trabajos de Buckland y Gey (Buckland y Gey, 1994) y también en Fugmann (Fugman, 1993, 1994). Por tanto, las relaciones semánticas, emanadas de la forma que aquí se ha expuesto y entendidas como la materialización de una estructura que permite la polirrepresentación conceptual, pueden ser un elemento importante para elevar la exhaus-

tividad de la recuperación de la información sin deterioro de la pertinencia, lo que de nuevo está en sintonía con los últimos avances teóricos en el campo de la Documentación, al tiempo que pueden ser una respuesta a parte de los problemas que tienen actualmente planteados los sistemas de recuperación de la información.

Notas

- (1) La importancia y la actualidad del aspecto semántico y de sus materias relacionadas puede fácilmente verse en un reciente artículo sobre la relevancia, publicado en ARIST (Schamber, 1994). Este artículo, a su vez, nos asoma a una amplísima bibliografía que sirve para apoyar el interés y el peso de las cuestiones semánticas en la Documentación.

Bibliografía

- Blair, D.C. (1986). Full text retrieval: evaluations and implications. // *International classification*. 13 : 1 (1986) 18-23.
- Blair, D.C. ; Maaron, M.E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. // *Communications of ACM*. 28 : 13 (1985) 289-299.
- Encabezamientos de materia de la Organización de Estados Americanos (1985). Bogotá : 1985.
- Fugmann, R. (1994). Galileo and the inverse precisio/recall relationship. Medieval attitude in modern Information Science. // *Knowledge Organization*. 21 : 3 1 (1994) 153-154.
- Fugmann, R. (1993). Subject analysis and indexing : Theoretical foundation and practical advice. Frankfurt : Indeks Verlag, 1993.
- Gardin, J.C. (1979) . On the relation between question-answering systems and various theoretical approaches to the analysis of text. // Maccaferry, Maxine y Gray, Kathleen (eds.). *The analysis of meaning. Informatics 5. Proceedings of a Conference held by the Aslib Informatics Group and the BCS Information Specialist Group, Oxford, 1979*. London : Aslib, 1979. p. 207-208.
- Gilchrist, A. (1971). *Thesaurus in retrieval*. London, Aslib, 1971.
- Horland, B. ; Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. // *Journal of the American Society for Information Science*. 46 : 6 (1995) 400-425.
- Ingwersen, Peter (1992). *Information retrieval interaction*. Los Ángeles : Taylor Graham, 1992. p. 50-56.
- Lancaster, F.W. ; [et al.] (1991). Identifying barriers to effective subject access in library catalogs. // *Library Resources and Technical Services*. 35 : 4 (1991) 377-391.
- Library of Congress Subject Headings (1991). Washington, 1991.
- López-Huertas, María J. (1997). Thesaurus structure design: a conceptual approach for improved interaction.// *Journal of Documentation*. 53 : 2 (1997) 139-177.

- López-Huertas, María J. (1998). Sobre el diseño de tesauros: algunas reflexiones teóricas y metodológicas. // II Encuentros sobre Organización del Conocimiento en Sistemas Documentales. Madrid : Isko-Capítulo Español, 1998.
- Maaron, M.E. (1988). Probabilistics design principles for conventional and full-text retrieval systems. // *Information Processing and Management*. 24 : 2 (1988) 249-255.
- Robertson, S.E. (1979). Between analysis and meaning. // Maccafferty, Maxine ; Gray, Kathleen. *The analysis of meaning : Informatics 5 : Proceedings of a Conference held by the Aslib Informatics Group and the BCS Information Specialist Group, Oxford, 1979*. London : Aslib, 1979. p. 203.
- Saracevic, Tefko (1975). Relevance : A review of and a framework for the thinking on the notion in Information Science. // *Journal of the American Society for Information Science*. 26 : 6 (1975) 321-343.
- Schamber, Linda. Relevance and information behavior. // *Annual Review of Information Science and Technology*. 29 (1994) 3-47.