

# Construcción de un tesauro para el Centro de Documentación de Telecinco

**Redacción:** Félix del Valle Gastaminza, Universidad Complutense de Madrid; Antonio García Jiménez, Universidad Rey Juan Carlos I

**Equipo de trabajo:** Inmaculada Chacón, José Ramón Laiseca, Cristina de la Vega, María Eugenia Martín Giraldo, Blanca Gil Urdiciain, Antonio Hernández, David Rodríguez

## 0.1. Resumen

Se analiza el estudio, la construcción y la puesta en funcionamiento de un tesauro en el entorno de la información audiovisual. Se analizan los descriptores para establecer su control y uso normalizado y se analizan de forma crítica las relaciones jerárquicas y asociativas propuestas en el marco de la macroestructura facetada y la estructura temática diseñadas a partir de los presupuestos de la teoría de los lenguajes epistemológicos. Se describen finalmente las etapas en la realización del trabajo.

**Palabras clave:** Tesauros. Construcción. Medios de comunicación. Documentación audiovisual. Televisión. Telecinco.

## 0.2. Abstract

A study about the design, construction and working of a thesaurus in an audiovisual context is presented. Firstly, the keywords, a very important part of the thesaurus, are analysed in order to study standardisation and connotation. Secondly, hierarchical and associative relationships are explained in relation to their diversity and their ability to improve the global structure in the thesaurus. Thirdly, the conceptual framework of the thesaurus is also analysed. The aim is to overcome thematic categorisation in favour of other textual models. In reality, this project is realized both from a traditional perspective and from new ones that try to be objective and to improve processes and methods related to this domain. Finally, different phases in the work are reviewed.

**Keywords:** Audiovisual archives. Television Archives. Thesauri.

## 1. Introducción: La documentación en Telecinco

El *Tesoro para materiales audiovisuales de Telecinco* se construye en respuesta a una iniciativa de los Centros de Documentación de esta cadena de televisión ante la necesidad de integrar su documentación en un entorno digital muy innovador y ante la constatación de la necesidad de disponer de una herramienta de lenguaje ágil y abierta capaz de tratar todo el conjunto de materiales generados por un sistema de producción de contenidos dirigido a la información y el entretenimiento.

La documentación en Telecinco se organiza en dos departamentos que hasta ahora habían resultado independientes entre sí hasta tal punto que sus normas, métodos, lenguajes, etc., eran diferentes aunque los usuarios y sus necesidades fueron similares. Por un lado, dependiendo de la División de Programación, desarrollaba su actividad el Centro de Documentación de Programas; por otro lado, dependiendo de la División de Informativos trabajaba el Centro de Documentación de Informativos, ambos ocupados de organizar todo el material audiovisual de la cadena.

El centro de documentación de programas se dedica, fundamentalmente, a procesar, analizar y conservar los programas de producción propia susceptibles de análisis. Por ejemplo, en la actualidad se tratan, entre otros, los siguientes: *Día a día*, con María Teresa Campos, *Caiga quien caiga*, *Nosóломúsica*, *Crónicas Marcianas*, *La gran ilusión*, etc. Su trabajo consiste en determinar las imágenes que puedan resultar más relevantes e interesantes para los futuros usuarios de los documentos, tanto los de la propia Telecinco como los ajenos a la cadena: Se presta especial atención a las personas que aparecen en entrevistas y actuaciones; los temas que se tratan, como los del corazón y de actualidad; los lugares que se encuentran en los documentos; las fechas en que fueron emitidos determinados programas; la trama de la que trata una película, etc. Este trabajo permitirá que los usuarios en su momento puedan explorar entre la información contenida en los más de cuarenta campos de la base de datos y especialmente entre los 13.500 documentos —entre programas, brutos y fichas principales— analizados hasta la actualidad con la posibilidad de encontrar con facilidad la información o las imágenes deseadas.

Para unificar la entrada de todos estos datos, de cara a una correcta recuperación, se ha utilizado un extenso manual de normas, así como otras muchas fuentes totalmente normalizadas. Otro trabajo que se realiza en el área de Documentación de Programas es la confección de las denominadas “fichas principales”, consistentes en unos registros en los que se incluyen los datos genéricos relativos a cada serie o programa emitido por la cadena: número total de capítulos de un programa o serie, periodicidad de su emisión, cuadro técnico-artístico, género, etc. Finalmente, este trabajo no tendría ningún sentido si no se le diera la debida difu-

sión, para lo cual todo el equipo de Documentación de Programas está siempre pendiente de atender y resolver las consultas que los diferentes usuarios hacen sobre los documentos. En definitiva, su trabajo consiste en determinar las imágenes que puedan resultar más interesantes para los usuarios.

Principios similares enmarcan el trabajo del Centro de Documentación de Informativos, pero la diferencia está en que este centro se ocupa exclusivamente del material de información de actualidad: los materiales brutos y emitidos que constituyen las noticias. Su trabajo se desarrolla en contacto directo con sus propios usuarios y su integración en la redacción es muy completa.

La creación del archivo digital diseñado para ser el núcleo del proceso de producción de televisión ha obligado, sin embargo, a replantear las cosas, pues los documentalistas asumen un papel muy activo y de enorme responsabilidad en la definición del archivo.

## 2. Análisis documental de materiales de televisión

La operación de indización y representación de material de informativos de TV, realizada con el objeto de hacer recuperable toda la documentación conservada, sigue los pasos siguientes:

- *Segmentación jerárquica del video en segmentos, secuencias y planos*: Un plano es una toma continua de cámara. Una secuencia está compuesta por uno o más planos que presentan diferentes tomas del mismo acontecimiento con una relación espacial y temporal. Un segmento está compuesto por una o más secuencias relacionadas.
- *Descripción de la información videobibliográfica completa*: título/s, autor/es, fechas, datos de publicación/edición, formato, duración, etc.
- *Descripción de cada segmento*: TCR (*Time Code Recording*) del *frame* de comienzo, duración, descripción textual, indización.
- *Descripción de cada secuencia*: TCR (*Time Code Recording*) del *frame* de comienzo, duración, descripción textual, indización.
- *Descripción de cada plano*: TCR (*Time Code Recording*) del *frame* de comienzo, duración, descripción textual, indización.

Para realizar todas estas operaciones el documentalista tiene que visionar el documento completo a veces más de una vez, tomando notas y sintetizando el contenido sonoro y visual para transformarlo en texto. Tiene que atender a la esencia del contenido y detectar cómo se ha expuesto en el documento; localizar los temas, personas y lugares relevantes; observar los aspectos formales y compositivos reseñables. El resultado final se expone en un registro catalográfico de una base de datos integrado por campos especialmente diseñados para los documentos en cuestión.

En el centro de documentación de Telecinco se está comenzando a utilizar nuevos programas y tecnologías que permiten automatizar algunas de las operaciones descritas hasta aquí y plantear nuevas formas de representación. Estas tecnologías utilizan técnicas avanzadas de análisis de imagen y sonido que son capaces de comprender automáticamente la señal de video. Detectan cambios de plano y movimientos de cámara, y generan un *storyboard* de fotogramas (*keyframe images*). Simultáneamente, extraen cualquier texto presente en la señal de video. Transforman la señal de audio en texto e identifican así palabras, hablantes, etc. Toda la información que genera es susceptible de contrastarse con un lenguaje documental incluido en el sistema y, por supuesto, es susceptible de revisión y ampliación por un documentalista. Estos programas no generan ya un solo registro por documento, sino que aprovechan la segmentación y jerarquización propia del documento para obtener un registro matriz del que dependen otros miniregistros con sus correspondientes “metadatos”.

El análisis documental de contenido de informativos audiovisuales da lugar a tres productos diferentes:

- El *resumen* del documento, de carácter *indicativo*, que define en pocas frases de qué trata.
- El *minutado*, que describe y mide con precisión lo que sucede en el documento, traduciendo a frases la dimensión temporal del contenido, e indicando la duración y contenido de cada escena que vendrá precedida por su TCR, lo que permitirá su localización automática.
- Los *descriptores*, resultado de la indización, que expresan el contenido en un lenguaje controlado.

Al realizar el visionado del documento el analista debe ir configurando esos tres productos, y para ello habrá de atender a una serie de atributos propios de la imagen como documento:

- *Atributos biográficos*: el autor o autores y, en general, aquellas personas que tienen alguna responsabilidad respecto al contenido, el tiempo y lugar de creación, el tiempo y lugar de emisión, el título, la serie a la que pertenece, los derechos de autor, etc...
- *Atributos temáticos*: de qué trata y cómo se trata — temas, personas, lugares, etc.—; qué, quién aparece en la imagen; cómo es lo que aparece; qué acciones se detectan, etc. Toda la riqueza audiovisual del documento habrá de ser representada en un nuevo lenguaje; éste es quizás el gran reto de la documentación audiovisual
- *Atributos relacionales*: Las imágenes de televisión pueden estar relacionadas o asociadas con otras imágenes o con textos y es importante descubrir

y tener en cuenta el contenido y la naturaleza de esas relaciones, que pueden resultar muy útiles para el usuario.

- *Atributos morfológicos*: Procedencia de las imágenes, tipos de planos utilizados, escenas especialmente significativas, puntos de vista (picado, contrapicado), movimientos de cámara (travelling, zoom, panorámicas), inclusión de gráficos, fotografías, mapas, sonido natural, locución, utilización del blanco y negro, etc.

El resultado final del análisis será una representación del documento que pasará a formar parte del fichero automatizado, de la base de datos, siendo, a partir de entonces recuperable. El análisis documental de contenido se diversifica y nos ofrece, en esta ficha, diferentes resultados que cumplen la función de informar sobre el contenido, pero, sobre todo, de garantizar su recuperación. Este documento podría ser recuperado por cualquiera de los descriptores que contiene, o por cualquier combinación de ellos, pero también se podría recuperar utilizando el lenguaje natural libre por cualquier palabra, serie de palabras o combinación de palabras en el resumen o el minutado.

Los documentos audiovisuales se caracterizan por estar integrados por dos planos de contenido que requieren un tratamiento diferente: Se trata de imágenes “de” algo que tratan “sobre” algo. La distinción entre “de” y “sobre”, entre referencia y connotación, se plasmará en dos tipos de descriptores organizados en tres grupos:

- *Lugar/plano y lugar/materia*: Se utilizan para indicar los descriptores relativos a lugares (ciudades, barrios, calles) que aparecen realmente en la imagen (lugar/plano) y para indicar los lugares donde se sitúa la noticia pero no aparecen en la imagen (lugar/materia). Por ejemplo, una imagen de un atasco de tráfico en una calle de Madrid llevará el descriptor /MADRID/ en el campo de Lugar/Plano; una entrevista con el Concejal de Tráfico del Ayuntamiento de Madrid en la que habla sobre el tema sin que aparezcan imágenes llevará el mismo descriptor en el campo de Lugar/Materia.
- *Tema/plano y tema/materia*: Situamos en estos campos los descriptores relativos a temas (entes concretos o abstractos) tratados en el documento, con las mismas características que en el caso de los descriptores geográficos. En el ejemplo citado el descriptor /TRAFICO/ se sitúa en el campo de Tema/Plano en el primer caso y en el de Tema/Materia en el segundo.
- *Persona/plano y persona/materia*: Campos para descriptores onomásticos (personas físicas y personas jurídicas). Una entrevista con Roberto Carlos en la que realiza declaraciones sobre Rivaldo y el Barcelona sería indizada con los descriptores /ROBERTO CARLOS/ en el campo de Plano/persona y con los descriptores /RIVALDO/ /BARCELONA FUTBOL CLUB/ en el campo de Persona/Materia.

- *Identificador/plano e identificador/materia*: Este campo se utiliza para indizar temas informativos de largo alcance sobre los que se genera mucha información. No suelen ser descriptores normalizados. Por ejemplo, podrían ser identificadores utilizados recientemente: /CASO GESCARTERA/ /ELECCIONES GALICIA 2001/ /GUERRA DE YUGOSLAVIA/.

En este entorno de trabajo ya se había percibido que la normalización del lenguaje era absolutamente necesaria y se habían puesto en práctica iniciativas diferentes. En el Centro de Documentación de Programas se había creado un lenguaje documental en el que fundamentalmente se habían controlado sinónimos y términos muy específicos que se trataban como sinónimos. También se habían desarrollado varias listas de vocabulario. Por su parte, en Informativos se había tratado de controlar el vocabulario y, para ello, se había generado un listado de descriptores válidos que, en la medida de lo posible, dado el número de trabajadores y turnos, era utilizado para la indización. En este contexto y bajo estas condiciones se acomete la tarea de diseñar y poner en marcha un Tesauro para integrar los materiales audiovisuales de Telecinco.

### **3. Los lenguajes epistemográficos**

En el principio de nuestro trabajo está una línea de trabajo: la propuesta de los nuevos lenguajes epistemográficos, que representan un intento de solucionar tanto el anquilosamiento de los lenguajes documentales tradicionales como su falta de adecuación al campo de la información de actualidad y al dominio de los materiales audiovisuales. En este sentido, no se debe perder de vista la tendencia que hace hincapié en nuevos desarrollos de este tipo de herramientas, tales como las ontologías, en la medida que son modos formalizados de representación de conocimiento para su recuperación.

Los lenguajes epistemográficos (García Gutiérrez, 1998), contruidos a partir de métodos científicos, son construcciones cognitivas que funcionan como una base de conocimiento con fines documentales. Su punto de partida es la crítica a las normas de construcción de los tesauros tradicionales, anclados en la infrutilización de las relaciones asociativas, en la falta de modelos que incorporen el contexto discursivo y el contexto de uso, y en la falta de conexión con las posibilidades que los nuevos programas informáticos ofrecen. Su potencia procede de su capacidad para crear representaciones cognitivas derivadas de desarrollos lógico-semánticos y discursivos que tienen como objetivo plantear representaciones pragmáticas (referidas al uso) de mapas cognitivos, simulando enunciados del discurso mediante la macroestructura y la microestructura. En realidad, se trata de lenguajes de estructura asociativa que permiten normalizar el vocabulario y que sugieren alternativas de búsqueda de información.

Los cambios respecto a los lenguajes documentales clásicos se producen en tres niveles. En primer lugar, la base léxica se intenta aproximar más a la *lengua natural*, mediante el uso de adjetivos, adverbios, prefijos, desinencias y verbos (en gerundio y en participio). El sustantivo, aunque se mantiene como la categoría léxica más importante, puede ser sustituido siempre y cuando suponga alguna modificación en el significado o conlleve la exclusión de otras categorías que proporcionen mayores prestaciones. De esta manera, se incluyen los adjetivos como descriptores calificadores del lenguaje y los verbos (mediante el gerundio y el participio) para representar las acciones, sin desdeñar otros elementos léxicos como es el caso de los prefijos, etc.

En segundo lugar, también se plantean modificaciones en las *reglas morfosemánticas* por las que se estructura el lenguaje documental, tanto en su construcción como en su uso. En el caso de los lenguajes epistemográficos, existen un mayor número de opciones en composición morfológica, y el número y el género adquieren más valor en función de los componentes significativos que proporcionan. Por otra parte, el usuario puede acceder a cualquier definición y nota de alcance, al mismo tiempo que elige la entrada que entiende más interesante para cada concepto.

En tercer lugar, con respecto a la estructura, cabe subrayar la *preponderancia de las relaciones asociativas* en detrimento de las jerárquicas, que desaparecerán una vez construido el vocabulario, figurando sólo a través de los operadores de generalidad (TG) y especificidad (TE). Concretando más, potencian el nivel sintagmático y la articulación del lenguaje documental, de tal modo que la jerarquización pierde su preponderancia en favor de las construcciones asociativas, si bien mantiene funciones de ayuda a la construcción. En efecto, las jerarquizaciones se mantienen, aunque en forma de encadenado ascendente (con género y todo) y encadenado descendente (con relaciones de clase, parte y enumeración), sin que suponga desdeñar su valor en tareas relacionadas con la detección de polisemias y de facilitación de las relaciones horizontales sobre campos de mayor homogeneidad y simetría.

Otro nuevo elemento que surge al amparo de esta propuesta es el , en el que gracias a los nuevos componentes tecnológicos se puede mostrar, en términos generales, las relaciones inmediatas (desde un punto de vista semántico, independiente de su profundidad) de un descriptor dentro de un sistema conceptual. En realidad, aunque siempre se manifiestan en función del usuario, hay que hacer referencia a dos tipos de escenarios, los intracategoriales o coordinados y los intercategoriales o asociativos.

A raíz de este planteamiento, que va dejando paulatinamente de ser teórico, dos son los elementos que merecen ser analizados con más detenimiento, a saber: las macrocategorías y las relaciones que se establecen entre descriptores, en la

medida que, a nuestro juicio, se convierten en las piedras angulares de este tipo de instrumentos. Las macrocategorías tienen un origen temático, empírico y aplicado, y cumplen con la función de organizar los campos conceptuales superando la adscripción temática. Suponen el primer nivel de la macroestructura global y son capaces de organizar los términos desde ópticas universales abstractas que no deben representar de forma directa los conceptos y los objetos reales del discurso que organiza, y que además tienen su evidente unión con las propuestas facitarias en la tradición documentológica. Tienen un procedencia cuando menos empírica y aplicada; y, para cumplir sus objetivos, deben cumplir los siguientes principios:

- *Inmutabilidad*: las denominaciones y los propios campos semánticos no pueden ser modificados en función de su situación lingüística y discursiva.
- *Exclusividad*: en general, cada macrocategoría no puede solaparse con otra, salvo en algunas situaciones operativas que deben ser absolutamente controladas.
- *Exhaustividad*: el conjunto de macrocategorías deben agrupar todos los términos que se incluyen en el lenguaje
- *Gramaticalidad*: la construcción de un sistema macrocategorial debe también estar orientado a la posterior propuesta de relaciones sintagmáticas.
- *Aplicabilidad*: en general, las macrocategorías se generan conforme a procedimientos empíricos para permitir tanto la representación discursiva como las demandas informativas.
- *Virtualidad*: la plasmación de las macrocategorías no se sitúan en el plano de trabajo del usuario sino que sirven para organizar y modelizar el discurso terminológico.

Por otra parte, en los lenguajes epistemográficos se establecen dos tipos de relaciones entre descriptores: las coordinaciones o relaciones intracategoriales, que mediante vectores internos establecen relaciones entre términos pertenecientes a la misma categoría, y que tienen su razón en la amplitud de los campos categoriales, y las asociaciones o relaciones intercategoriales, que en realidad son relaciones asociativas determinadas por vectores asociativos que presentan la dirección de las relaciones entre términos de distintas macrocategorías. Estas relaciones requieren una descripción de las categorías que se realice de modo excluyente, una adscripción excluyente de los conceptos en las categorías y asimismo un texto que defina el comportamiento del descriptor en el mismo. En definitiva, el vector —que es un elemento teórico más que práctico que procede de la forma de comprensión del discurso correspondiente, y que resulta de la combinación de elementos macrocategoriales en un entorno microestructural— se convierte, al concretar las relaciones entre los descriptores, en el elemento básico de la microestructura, entendida como una construcción relacional mínima establecida por



dos elementos del lenguaje. Finalmente, hay que señalar que el desarrollo teórico y práctico ha estado vinculado en mayor medida a los vectores asociativos y menos a las coordinaciones o relaciones intracategoriales y esto se ha de solventar en epistemografías concretas, como es el caso que aquí se presenta.

Por otra parte se estudiaron con detenimiento las experiencias conocidas de lenguajes documentales realizados para el tratamiento del contenido de las imágenes, notoriamente el *Art & Architecture Thesaurus* creado por la fundación Paul Getty Trust cuya estructura facetada está muy próxima a la idea de macroestructura que proponemos para el tesoro de Telecinco.

#### **4. Construcción del tesoro**

Este tesoro se perfila como un sistema y un instrumento lingüístico con vocación temática y paradigmática, que presenta una estructura básica conformada por un conjunto de términos descriptores controlados que deberán representar o describir los términos, conceptos e ideas susceptibles de aparecer en televisión.

Se trata pues de un área temática y contextual de muy largo alcance, casi enciclopédica, caracterizada por una enorme dispersión semántica y una gran variedad de nivel de ponderación de descriptores en donde era imprescindible actuar en favor de una mayor conexión con el discurso con el que se trabaja y con las necesidades de los usuarios.

El punto de partida material, ya citado, han sido los listados de descriptores controlados propuestos por cada centro de documentación, listados alfabéticos de gran dispersión temática como lo demuestran descriptores como: /ACTRICES/ /ALCANTARILLADO/ /ALIJOS/ /ARROZ/ /EXILIO/ /ERRORES MEDICOS/ /OSOS PANDA/ /SIAMESES/ /SUBSAHARIANOS/ o /SECRETARIAS/.

En definitiva, se ha tenido que analizar la base léxica en un marco de trabajo unido a la información audiovisual, donde adquiere mayor relevancia la connotación a la hora de representar contenidos y significados.

En segundo lugar, se analizan las cuestiones que plantean las relaciones jerárquicas y asociativas, así como las relaciones de equivalencia, desde la perspectiva de su construcción, de su diversidad y diferenciación tipológica, de su profusión y de su capacidad para mejorar la estructura global del tesoro.

En tercer lugar, sin apartar las dificultades que surgen en la compilación terminológica del corpus alfabético, se aborda la determinación del esquema conceptual que sirve de estructura al sistema lingüístico que se construye. En esta ocasión, se opta por una estructuración global que, por un lado, pretende superar la simple adscripción temática por parte de los descriptores en beneficio de otros ángulos de estructuración conectados con el discurso de actualidad y que, por otro lado, se acerque más a lo empírico que a lo especulativo.

#### 4.1. Macroestructura facetada, estructura temática y escenario

La construcción del tesoro parte de la definición de la estructura. En este caso, se realizará sobre dos ejes: un eje vertical que estructura jerárquicamente el tesoro en torno a un conjunto predeterminado (y modificado de modo empírico posteriormente) de macrocategorías, y un eje horizontal que cruza transversalmente todas las facetas integrado por las áreas temáticas cubiertas por la televisión. Las macrocategorías provisionales y las áreas temáticas aparecen al final de las fases. Si bien el procedimiento teórico adecuado para la creación del conjunto de macrocategorías es hacerlo depender de los términos encontrados o definidos, por motivos operativos, en esta ocasión, la propuesta de macrocategorías que se establece para la configuración de la macroestructura tiene su origen en la aplicación conjunta de propuestas facetarias, en principio independientes del discurso en el que nos situamos, procedentes de diversas fuentes como el *Art & Architecture Thesaurus*, García Gutiérrez, Ranganathan y van Slype. El planteamiento inicial es el siguiente:

- a) *Fenómeno*: acción natural que escapa a la acción del ser humano y acontecimientos no provocados por el ser humano; por ejemplo, digestión, retraso.
- b) *Actividad*: acción provocada por el ser humano.
  - b1) *Acontecimiento*: denominación genérica de sucesos provocados por el hombre de gran relevancia o resonancia que afectan directa o indirectamente a los seres humanos.
  - b2) *Funciones*: campo de la actividad profesional
  - b3) *Disciplina*: rama del conocimiento, arte, ciencia, tecnología.
  - b4) *Técnicas*
  - b5) *Actividad física*: por ejemplo, un deporte concreto.
- c) *Materiales*: sustancias naturales o producidas artificialmente, incluyendo energía.
- d) *Agente*: Describe personas, seres vivos y grupos de ambos por su actividad, sus características físicas y culturales, rol o condiciones social.
  - d1) *Colectivo*: instituciones, organizaciones o colectivos humanos o con representación humana. En general, entidades complejas creadas por seres vivos.
  - d2) *Individual*: seres humanos, (con sus profesiones, roles y ocupaciones personales), microorganismos, vegetales y animales.
- e) *Objeto*: cosas inanimadas, visibles o tangibles resultado de la actividad humana.

- e1) *Objeto inmueble*: espacios y construcciones y sus clases, partes, componentes y dependencias; por ejemplo, valla, ciudad, jardín, edificio inteligente
- e2) *Objeto mueble*: cosas materiales, instrumentos y artefactos; agrupaciones de objetos y sistemas, componentes y géneros de objetos según su forma y función, formatos de comunicación y documentos. Ejemplos: sistema de comunicación, fragmento, antena parabólica, vehículo, mural, examen, catapulta.
- f) *Estructura*: conceptos teóricos, abstracciones, elementos intangibles estáticos y teorías que componen o describen los sistemas socioeconómicos, sociopolíticos e ideológico.
- g) *Atributos*: características, cualidades y propiedades de agentes, acciones, objetos, materiales y estructuras.

Los temas provisionales del tesoro fueron extraídos de los códigos propuestos por el International Press and Telecommunications Council (IPTC) para la transferencia internacional de noticias. Las áreas temáticas son las siguientes:

- Arte. Cultura. Espectáculos.
- Justicia.
- Desastres y accidentes.
- Economía. Negocios. Finanzas.
- Educación.
- Medio ambiente.
- Salud.
- Trabajo.
- Ocio. Estilo de vida.
- Política.
- Religión. Creencias.
- Ciencia y tecnología.
- Asuntos sociales.
- Deporte.
- Guerras. Conflictos.

En cuanto a las relaciones propuestas para cada descriptor, el punto de partida ha sido el siguiente:

- *Relación de jerarquía* (incluye todo/parte y género/clase): entre un descriptor y cada uno de los elementos jerárquicos señalados.

- *Relación actancial*: entre un descriptor y el actante, que es quien puede ejecutar una acción o sufrirla.
- *Relación modal* (incluye acontecimientos, actividades, procesos y técnicas): relación de un descriptor y un acontecimiento, actividad, proceso o técnica en el sentido que antes se ha descrito.
- *Relación de atributo*: relación entre un descriptor y las características, cualidades y propiedades de agentes, acciones, objetos, materiales y estructuras.
- *Relación estructural*: relación entre un descriptor y los conceptos teóricos, abstracciones, elementos intangibles estáticos y teorías que componen o describen los sistemas socioeconómicos, sociopolíticos e ideológicos.
- *Relación de materia/causal*: relación entre un descriptor y sustancias naturales o producidas artificialmente, incluyendo la energía.
- *Relación de objetos/productos* (incluye objetos inmuebles, objetos muebles, instrumentos o materiales transformados): relación entre un descriptor y los espacios y construcciones y sus clases, partes, componentes y dependencias, o bien entre un descriptor y las cosas materiales, instrumentos y artefactos, en los que también se incluyen las agrupaciones de objetos y sistemas, los componentes y géneros de objetos según su forma y función, así como los formatos de comunicación y documentos.
- *Relación locativa de espacio*: en este caso esta propuesta teórica se ha subsumido en la anterior relación.

Con respecto a la norma UNE, la relación TR quedaría de la siguiente forma (variantes):

- *Una disciplina o campo de estudio y los objetos o fenómenos estudiados*: Plantea ciertas dudas a la hora de incluirla en la tipología de relaciones propuesta.
- *Una operación o proceso y su agente*: Se incluye mediante la relación actancial y la modal.
- *Una acción y resultado o producto de tal acción*: Se puede incluir en otras categorías de relaciones de objetos/productos
- *Una acción y su sujeto pasivo*: Se incluye mediante la relación actancial y la modal.
- *Conceptos y sus propiedades*: Se incluye en la relación de atributo.
- *Conceptos relacionados con sus orígenes*: Se incluye en la relación de materia causal.
- *Conceptos ligados por una dependencia causal*: Se incluye en la relación materia/causal.

- *Objetos y sus agentes*: Se incluye en la relación actancial.
- *Conceptos y sus unidades de medida*: Se puede incluir en instrumento

Por lo que, finalmente, el escenario propuesto para cada descriptor es el siguiente: descriptor, macrocategoría o faceta, campo conceptual o temático (código), nota aclaratoria o de aplicación, equivalencias (UP), observaciones, relaciones asociativas, relaciones de coordinación.

#### 4.2. Etapas en la elaboración del tesoro

A partir de los puntos teóricos de interés, anteriormente reseñados, se establecieron los siguientes procedimientos y etapas de trabajo:

1. *Incidencias en la obtención de los términos candidatos* a formar parte del tesoro, teniendo en consideración la experiencia previa del centro de documentación sobre el que se trabaja y que ha recopilado, a lo largo de su proceso natural de trabajo, unos 2.000 descriptores. Asimismo, se hace referencia a la importancia y relevancia del uso de otros lenguajes documentales (terminologías, clasificaciones, diccionarios, tesauros) relacionados con la información audiovisual de actualidad susceptibles de aportar términos al tesoro.
2. *Definición de la estructura*: Tal y como se ha comentado, se realiza sobre dos ejes: un eje vertical que estructura jerárquicamente el tesoro en torno a los casos gramaticales, especialmente adecuados para tratar contenidos, y un eje horizontal que cruza transversalmente todas las facetas integrado por las áreas temáticas cubiertas por la televisión.
3. *Ubicación de los descriptores en cada área*: Todos los miembros del equipo trabajaron con cada uno de los 2.000 descriptores propuestos inicialmente para asignarlos a una o varias áreas temáticas. Esto resultó de enorme utilidad para familiarizarse con todo el vocabulario. Un porcentaje muy alto de descriptores estaba en más de una de las áreas
4. *Ubicación de los descriptores en cada faceta*: Igualmente todos los miembros del equipo trabajaron sobre cada descriptor. En este caso era obligatorio asignar cada descriptor a una sola faceta.
5. *Control de sinónimos*: En la medida en que cada concepto debe estar representado por un descriptor o combinación de descriptores.
6. *Establecimiento de relaciones jerárquicas dentro de cada macrocategoría*: Nivelación de facetas. Al realizar este trabajo se han observado muchas carencias en el tesoro y muchos problemas derivados del trabajo con un simple listado. Se han completado las facetas y para ello se ha recurrido a muchas fuentes, entre las que es obligado destacar el léxico para

materiales informativos de Lexis-Nexis, el Macrotesauro de la OCDE, El tesauro Eurovoc, el tesauro de Población de Unesco y fuentes directas como el Código Penal o los Códigos de Disciplinas y Ciencias de Unesco. La investigación de las facetas puso al descubierto muchos problemas estructurales: Así a modo de ejemplo pueden verse los casos siguientes:

- En la terminología relacionada con el *deporte* se encontró con que aparecían descriptores relativos a los *agentes deportivos* (como /FUTBOLISTAS/ /CICLISTAS/ o /TENISTAS/), otros relativos a las *actividades deportivas*, como /FUTBOL/ /CICLISMO/ o /TENIS/ aunque ni mucho menos estaban todos los deportes ni todos los deportistas. La propuesta final ha sido suprimir a los agentes para sustituirlos por la poscoordinación entre el descriptor relativo al deporte y otros relativos a agentes, como /DEPORTISTA/ /JUGADOR/ o /ENTRENADOR/.
  - Algo similar, aunque a cuatro bandas no siempre completas ni equilibradas se ha dado entre las *enfermedades*, por ejemplo, /ALERGIA/, los enfermos /ALERGICOS/, los especialistas médicos /ALERGOLOGOS/ y, finalmente, su especialidad, /ALERGOLOGÍA/. Con esto no queremos decir que estuvieran los cuatro elementos para cada enfermedad, sino que para cada una aparecía una cosa diferente. Finalmente se propuso mantener la especialidad y la enfermedad que pueden combinarse con /MEDICOS/ y /ENFERMOS/.
  - Al provenir los descriptores de dos servicios diferentes se observó una discrepancia fundamental en los niveles de profundidad utilizados por unos y otros en distintas áreas temáticas. Discrepancia que, en principio, los documentalistas integrados en la construcción del tesauro percibían como un grave obstáculo y que, sin embargo, ha contribuido a dotar al tesauro de una mayor riqueza jerárquica.
  - También como consecuencia de esto se detectaron significados y connotaciones diferentes para un cierto número de descriptores, que hubo que solucionar mediante notas aclaratorias.
7. *Definición de las relaciones asociativas* utilizando la propuesta de relaciones vista arriba para establecer las relaciones de cada descriptor. Es la fase en la que estamos en este momento y para la que se ha propuesto una búsqueda individual exhaustiva por facetas que se perfilará en un trabajo colectivo de equilibrado.
  8. El *resultado final de facetas*, sin embargo, ha diferido de la propuesta original y ha quedado definitivamente como sigue:

- *Acontecimientos*: ACTIVIDADES. PROCESOS. Donde figuran descriptores como /ABASTECIMIENTO DE AGUA/ /AGRICULTURA/ /FIESTAS PATRONALES/ /GENOCIDIO/ /INSEMINACIÓN ARTIFICIAL/ /MANIFESTACIONES/.
  - *Agentes*: Con descriptores como /ABUELOS/ /ACROBATAS/ /AUDIENCIA NACIONAL/ /CONCEJALES/ /COOPERATIVAS AGRICOLAS/ /PSICOFONISTAS/.
  - *Objetos*: /ACCESORIOS DEL AUTOMOVIL/ /ALCANTARILLADO/ /ARMAS QUIMICAS/ /LAVAVAJILLAS/ /TARJETAS DE CREDITO/.
  - *Materiales*: /AFRODISIACOS/ /ALIMENTOS DE REGIMEN/ /COBRE/ /PERFUMES/ /RESIDUOS TOXICOS/.
  - *Estructuras*: /ANARQUISMO/ /AUTODETERMINACION/ /EMPLEO/ /FEMINISMO/ /RACISMO/ /VIOLENCIA/.
  - *Atributos*: Esta faceta tiene un desarrollo especial y, por un lado, integra los atributos propiamente dichos, descriptores que complementan a otros descriptores y que sólo se utilizan formando parte de un nuevo descriptor, como /PROFESIONAL/ o /PRIVADO/ y, por otro lado, se han organizado dos subfacetas muy importantes como son la de *acciones*, con descriptores como /SALTANDO/ o /DESPEGANDO/ y la de *expresiones faciales* que incluye descriptores como /ALEGRIA/ /ANSIEDAD/ o /SERIEDAD/
  - *Listados auxiliares*: Se ha desarrollado un listado de Anatomía y Biología, con descriptores como /COLESTEROL/ o /SISTEMA DIGESTIVO/; y un listado de Estilos artísticos, musicales y cinematográficos que incluye descriptores como /SURREALISMO/ /JAZZ/ /COMEDIA MUSICAL/ o /BULERÍAS/
9. *Edición del tesoro e integración en el archivo digital*: La edición inicial del tesoro va a ser impresa y se comenzará a utilizar en la indización de forma inmediata. Se propondrá un listado alfabético de descriptores con su correspondiente escenario de relaciones y con todas las notas de aplicación y situación en temas y facetas. Este listado incluirá los tres mil descriptores controlados y los cerca de mil no descriptores con su correspondiente reenvío. Respecto a la integración en el archivo digital, basado en una aplicación diseñada por Informix, que integra software de Virage y de Excalibur, para Telecinco, se está trabajando en una aplicación en XML que permita la consulta y navegación por el tesoro en el entorno de trabajo habitual de los usuarios y de los documentalistas.

10. Señalaremos finalmente que se propondrá a Telecinco un programa de mantenimiento periódico por el cual se realice una primera validación del funcionamiento del tesoro y un sistema controlado de incorporación de descriptores, absolutamente necesario en un campo tan imprevisible como la información de actualidad: ¿Algún tesoro de información de actualidad, para prensa, radio o televisión, incluía el descriptor Ántrax hasta hace unos pocos meses? Ahora, sin embargo, no estarían dispuestos a prescindir de él.

## 8. Bibliografía

- AENOR (1990). Norma UNE 50-106-90: Documentación: Directrices para el establecimiento y desarrollo de tesauros monolingües. Madrid: AENOR, 1990.
- Aitchison, Jean; Alan Gilchrist and David Bawden (2000). *Thesaurus construction and use: a practical manual*. 4th ed. Chicago: Fitzroy Dearborn Publishers, 2000.
- García Gutiérrez, Antonio (1998). *Principios de lenguaje epistemográfico: la representación del conocimiento sobre Patrimonio Histórico Andaluz*. Sevilla: Instituto Andaluz del Patrimonio Histórico, 1998.
- Slype, Georges van (1991). *Los lenguajes de indización: Concepción, construcción y utilización en los sistemas documentales*. Madrid; Salamanca: Fundación Germán Sánchez Ruipérez; Pirámide, 1991.