

La indización en la recuperación de la información

Elisa Carolina de Santos Canalejo

INTRODUCCIÓN

No cabe duda que la informatización de los archivos ha supuesto la aplicación de unas técnicas documentales que hasta hace pocos años estaban más extendidas en el ámbito de las bibliotecas y de los centros de documentación. Aplicación, que, por otra parte, debe tener en cuenta las características propias del *documento de archivo y los diferentes soportes*. De ahí el reto para los archiveros en la selección y uso de esas técnicas documentales de acuerdo con la clase de archivo y tipo de documento.

Asimismo, desde hace años, y también como consecuencia de la informatización, se ha planteado en los archivos la necesidad de establecer normas en las tareas archivísticas. Dicha normalización debería abarcar todas las tareas archivísticas realizadas en cada tipo de archivo, desde la organización e identificación de fondos, hasta la descripción, la localización topográfica, etc.

Por eso la normalización constituye uno de los temas que más interés ha despertado en los últimos tiempos, como demuestra la edición de la Norma ISAD(G) en 1993¹ y la ISAAR (CPF)² en 1995 bajo los auspicios del Consejo Internacional de Archivos y que están ahora en fase de análisis y de prueba en los archivos. La Norma ISAD(G) aspira a ser la Norma Internacional General de Descripción Archivística, y la ISAAR (CPF) constituye la Norma Internacional sobre los Encabezamientos Autorizados Archivísticos relativos a Entidades, Personas y Familias, consecuencia de la primera Norma como comprobaremos más adelante.

Pero es preciso destacar que la normalización va ligada al control del vocabulario si se desea la más pertinente recuperación de información y su intercambio. De ahí, la importancia de la normalización terminológica, que afecta tanto al vocabulario profesional propiamente dicho como a las propias tareas de la descripción archivística. En este sentido, el control terminológico procede de los lenguajes de indización y supone la selección de los términos que representan al documento en sí, a su contenido y a su contexto para proporcionar la calidad de información que se pide actualmente a los archivos, insertos en un sistema de información y por ello cada día más frecuentados por un público muy heterogéneo.

No podemos olvidar que un sistema de información debe fomentar la utilización de la información que contiene, y por eso el archivo como centro de información debe considerar tanto los usuarios habituales como los potenciales a quienes debe dirigirse, desarrollando las técnicas de tratamiento de la documentación para la recuperación de la información.

Ahora bien, es importante destacar que el empleo de las normas acompañadas por el de estos lenguajes documentales modifica sustancialmente la manera de trabajar en los archivos y hasta repercute en la propia organización de un sistema nacional de archivos.

LA INDIZACIÓN EN LOS ARCHIVOS: LOS NUEVOS RETOS

Los términos índice e indizar hasta hace poco tiempo se referían en el ámbito de los archivos a un instrumento de descripción y a la realización de ese instrumento, pues de todos es conocido que entre los tradicionales instrumentos de descripción de un archivo figuran los índices, referidos especialmente a términos onomásticos de unos fondos documentales. Pero los archiveros también sabemos que cualquier instrumento tradicional de descripción, como una guía, un catálogo, un inventario, no estaría completo si no esta acompañado por unos índices que recuviera los términos onomásticos, geográficos y de materias, respectivamente, o bien un índice general que los integrara a todos ellos.

Así pues, indizar, es decir, extraer los términos más relevantes de los documentos para elaborar los índices, ha sido siempre una de las tareas realizadas por el archivero.

Ahora bien, esa elaboración manual de los índices es individual, o lo que es lo mismo, de cada archivero, y responde sólo a la fase última de la descripción destacando aquellos términos más sobresalientes de la documentación, sin una normalización terminológica previa. Esos índices interesan, sobre todo, para localizar los documentos referidos a un asunto, un lugar geográfico o una persona a los que se refiere cada instrumento de descripción.

Sin embargo, diversos factores como la importancia que hoy día tiene la identificación de fondos ante el volumen de documentación que existe en los archivos y la consiguiente aplicación de técnicas documentales en ellos, paralelamente a su informatización, plantean al archivero un nuevo concepto de indización, que tenderá a ser cada vez menos de creación individual y más de creación colectiva, y que afectará a las distintas fases de la descripción propiamente dicha.

Pues, no cabe duda que si un archivo quiere ser un auténtico centro de información e integrante de un sistema de información debe atender muy preferentemente todos los elementos que sean puntos de acceso de la información que contiene. Y, además, debe aspirar a que la difusión de su información sea primero a nivel nacional y luego, internacional, y para ello, deberá cooperar en la elaboración de puntos de acceso comunes a varios archivos.

Por ello, el archivero precisa la utilización de los llamados *lenguajes documentales*, es decir, de aquellos que por procedimientos sintácticos convencionales representan el contenido de un documento o de un conjunto de documentos para facilitar la búsqueda y recuperación de la información. Influyen, al tiempo que se forman, en el análisis y la recuperación de la información. De una manera básica, pueden dividirse en lenguajes de estructura jerárquica, como son las clasificaciones, y de estructura combinatoria, como son los tesauros y los índices, que para los archivos ofrecen unas características específicas en relación con los reseñados en los manuales y tratados aplicados a bibliotecas y centros de documentación, manuales que todavía no tratan a los archivos como centros que necesiten de estos lenguajes. La indización es la clave de los lenguajes documentales, por eso, en algunos manuales se habla de lenguajes documentales o de indización. Son lenguajes normalizadores y, a su vez, normalizados.

De esta manera, nos es útil para *el documento de archivo* la definición dada por Chaumier sobre el análisis documental como *la operación o conjunto de operaciones realizadas para representar el contenido de un documento bajo una forma diferente de la presentación original con el fin de facilitar su consulta o recuperación*, pues se trata de facilitar la recuperación de información y de conocer el origen y contenido de los documentos.

Este análisis documental en la descripción archivística constaría de un análisis formal, aplicando las normas (ISO, ISAD(G), ISAAR(CPF,...) y un análisis de contenido tendente a la indización. En este sentido, debe recordarse que la ISO ha dictado una norma que de un modo general también debe tenerse en cuenta en la indización en los documentos de archivo: se trata de la Norma ISO 5963-1985 *Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection de termes d'indexation*³.

En esta norma se define la indización como la *acción de describir o de identificar un documento en lo referente a su contenido*, que consta de tres etapas principales:

- examen del documento y definición de su contenido,
- identificación y selección de los términos principales del contenido,
- selección de los términos de indización.

Se ha dicho que la descripción en un archivo debe ser exacta, suficiente y oportuna, y que la información obtenida debe ir de lo general a lo particular y debe ser pertinente, no repetida e interconexa. Por tanto, se deberá indizar en cada nivel de información y, así, se establecerán términos de indización a nivel de fondo, a nivel de serie y a nivel de documento. Los cuadros de clasificación, los tesauros y los índices permutados se convierten de este modo en los lenguajes de indización a la vez que en instrumentos de descripción de los archivos.

No obstante, se debe ser consciente que la aplicación de estos lenguajes en los archivos supone un esfuerzo, máxime cuando en las bibliotecas y centros de documentación la mayoría de los lenguajes de indización han sido elaborados de forma independiente, sin integrarse o interconectarse en una superestructura común⁴.

Y también es cierto, que todavía el usuario de un archivo no está familiarizado con estos nuevos instrumentos de descripción, y esto exige un mayor reto a los archiveros que deben emplear con mucha claridad esos lenguajes para que los sistemas informáticos reenvíe automáticamente a los términos de indización cualquier otro que pida el usuario.

La informatización de los archivos supone, por tanto, la creación de bases de datos que incluyen los puntos de acceso para recuperar la información, y la utilización de los programas informáticos para la creación, desarrollo, mantenimiento y actualización del vocabulario controlado, su visualización en línea y su edición impresa⁵.

Y hablamos de usuario en general, y no del investigador especializado en particular, porque cada vez acuden más a los archivos curiosos y particulares, sin olvidar la propia administración, en el caso de los archivos administrativos.

En definitiva, el archivero, si desea conseguir recuperar la información con la mayor exhaustividad, precisión y pertinencia posible, deberá tener en cuenta la necesidad de indizar con vocabulario controlado en las distintas operaciones archivísticas de la descripción, e incorporar esa terminología normalizada en un sistema informático tanto para la consulta del usuario como para otros archivos de ámbito nacional e internacional, con el fin de establecer para cada especialidad, como por ejemplo, la de las relaciones internacionales, una terminología común normalizada.

LA NORMALIZACIÓN DE LA TERMINOLOGÍA EN LOS ARCHIVOS

Y hablamos de indizar con vocabulario controlado, pues también la selección de términos de indización puede elaborarse a partir de los términos contenidos en los documentos. Sin embargo, si la terminología de los términos de la indización no está normalizada, es decir, si no obedece a unos criterios objetivos y establecidos de antemano, no tendría sentido esa tarea pues perdería su eficacia y se desperdiciaría el aprovechamiento informático en la recuperación y difusión de la información⁶. Pero, entonces, cabe preguntarnos, ¿quién fija esos términos? Es nuestra experiencia personal y opinión que según el tipo de archivo y del fondo documental de que se trate los responsables de la indización deberán ser distintos:

- o Si se trata de un archivo administrativo con fondos y series documentales abiertas, no cabe duda que el establecimiento de criterios para la elección de los términos de indización deberá ser ejecutado entre el archivero y el órgano productor de la documentación. De esa manera, el órgano productor participa en la elaboración del lenguaje de indización que, por tanto, no le será ya nunca ajeno y desconocido, y la comunicación con el archivo siempre será fluida, facilitándose así el buen funcionamiento del sistema de información de ese organismo o institución, y contribuyendo desde el archivo a la modernización de la administración⁷. Téngase en cuenta, además, que la misma denominación del órgano productor, y del órgano remitente en su caso, también se deben normalizar como puntos de acceso.

Incluso para los fondos y series documentales cerradas que un archivo administrativo pudiera conservar, sería conveniente que el archivero solicitara información de su organismo para recabar la mayor información posible y, así, acercar los términos de indización al lenguaje más usado y conocido de ese organismo e institución.

Pues no hay que olvidar que el lenguaje de indización en un archivo administrativo debe ser el más utilizado por su organismo, es decir, el más cercano al lenguaje natural empleado por sus funciones y competencias. La selección del vocabulario es por lo general *a priori*, es decir preestablecido antes de empezar a indizar.

- o En los archivos históricos, la selección del vocabulario lo realiza el archivero y es *a posteriori*. Es decir, el archivero elabora el lenguaje de indización a partir del estudio de la documentación. De ahí que sea muy importante considerar la posibilidad de que para fondos documentales comunes en archivos históricos, se trabaje en equipo. De este modo, el intercambio de información se facilitará considerablemente.

Asimismo, conviene tener en cuenta la influencia de los medios de comunicación social y de la historiografía cuando se está creando ese lenguaje de indización para que resulte lo menos artificial posible y se acerque a la demanda del usuario. Al fin y al cabo, en el proceso de indización no sólo se ha de conocer el documento para seleccionar los puntos de acceso, sino también el usuario.

Por otra parte, si en función del tipo de Archivo, la indización tiene distintos responsables, también el tipo de archivo condiciona el grado de implantación de un vocabulario controlado:

- o En un archivo administrativo, la puesta al día implica, por un lado, la introducción provisional de nuevos términos de indización, que podríamos denominar *términos candidatos*, pues según su relevancia y frecuencia de uso se integrarán o no

definitivamente en la indización y, por otro lado, la supresión o modificación de términos según el índice de pertinencia y profundidad. Así pues, en un archivo administrativo el vocabulario controlado crece y se modifica apoyándose en la observación de su uso. Se va creando en el archivo administrativo, por tanto, un lenguaje de indización con términos "históricos", correspondientes a los fondos históricos que pasarán, después, a integrarse en el vocabulario del archivo histórico, y con términos "vivos" de acuerdo con el propio funcionamiento del órgano productor. La indización aparecerá, por tanto, reflejada en las relaciones de entrega, sea cual fuere el plazo de conservación de la documentación⁸.

- o En un archivo histórico, los términos que se seleccionen para la indización deberán ser lo suficientemente relevantes, pertinentes y profundos para que su implantación sea lo más definitiva posible, pues hacen referencia a documentación histórica. Precisamente su vigencia hará posible el intercambio de vocabularios entre Archivos, como por ejemplo, en el caso de fondos de Organismos Internacionales, o de fondos de época colonial, entre archivos de la metrópoli y los de las antiguas colonias.

En definitiva, la indización en los archivos presenta distintas características según el tipo de archivo. La normalización terminológica, por su parte, deriva de la propia documentación y de las obras de referencia que se precisen, y afecta tanto a la denominación del órgano remitente y productor de la documentación como a los términos relativos a su descripción.

LA INDIZACIÓN EN LA IDENTIFICACIÓN

Ya en la primera fase de la descripción archivística, esto es, en la identificación de los fondos y series documentales, se debe normalizar el vocabulario y codificarlo para su informatización. De esta manera, cada término fijado referido al órgano productor ya constituye un punto de acceso de información. De esta premisa parte la Norma ISAAR (CPF) sobre los Encabezamientos Autorizados Archivísticos relativos a Entidades, PersonasyFamilias presentadas por la Comisión Ad Hoc de Normalización Descriptiva.

En su introducción se justifican esos encabezamientos de autoridad para que la documentación de un mismo productor conservada en archivos diferentes, o por el mismo productor, se identifiquen de un modo normalizado y se facilite, además, el intercambio de información, incluso en el ámbito internacional. La normalización afecta, según esta norma, por tanto, a entidades, personas o familias como órganos productores.

Asimismo, la Norma sirve no sólo para controlar la creación y empleo de los encabezamientos como puntos de acceso, sino también para la descripción de una entidad, persona o familia, es decir, para proporcionar cuanta información se disponga sobre ese órgano productor. De ahí, la conexión con la Norma ISAD(G), pues, entre los elementos de la descripción citados en esa norma, se señala en el Área de Contexto (3.2), el nombre del productor (3.2.1) y la historia institucional o biográfica (3.2.2), y se prevé la posibilidad de organizar esa información en ficheros de encabezamientos autorizados relacionándolos con este Área de Contexto.

Según la Norma ISAAR(CPF), cada encabezamiento autorizado está compuesto por tres áreas:

- o **1ª Área de Control de Autoridad**, compuesta por el código del archivo responsable que crea la entrada de autoridad; el tipo, es decir, si es una entidad, persona o familia; la entrada de autoridad, o lo que es lo mismo, el punto de acceso nombrado; la(s) entrada(s) paralela(s), que son las formas que aparecen en otras lenguas; la(s) forma(s) excluida(s) que reenvían las denominaciones excluidas a la entrada de autoridad. Es lo que se denominaría véase, o usado por un lenguaje de indización; y la entrada(s) relacionada(s), que enlaza la entrada de autoridad con otras. Es lo que se utilizaría con un véase también en un lenguaje de indización.
- o **2ª Área de Información**, que tiene como finalidad describir la entidad, persona o familia. Los elementos de esa descripción se señalan en estas normas:
- o Para la entidad: número oficial de registro en el país, nombres, fechas y lugares de existencia, sede, estatus jurídico, funciones, organigrama, relaciones y otra información significativa.
- o Para la persona y para la familia, los nombres, fechas y lugares de existencia, lugar de residencia, nacionalidad, funciones y actividades, árbol genealógico en el caso de la familia, relaciones y otra información significativa.
- o **3ª Área de Notas**, que abarca: la nota del archivero, que señala las fuentes consultadas y las utilidades para su actualización para el establecimiento de esa entrada de autoridad; las reglas o normas nacionales o internacionales; y las fechas de cuando se estableció la autoridad y cuando se revisó.

No cabe duda que la descripción de cada encabezamiento autorizado es muy detallada. Sin embargo, conviene que analicemos cada una de esas áreas reflexionando sobre algunas cuestiones:

Respecto al *Área de Control de Autoridad*, el campo de código exige que todos los archivos tengan uno. Esto implica, por un lado, que, en el ámbito nacional, se codifique alfabética y numéricamente cada archivo responsable de registrar estos puntos de acceso, y, por tanto, se designen los archivos responsables de llevar a cabo esta tarea, pues quizás sería conveniente que en un sistema nacional de archivos se crearan grupos de trabajo de archivos responsables de normalizar las entidades, otros para las familias y otros para las personas. De hecho, desde la Subdirección General de Archivos Estatales éste es nuestro planteamiento. Se trata pues de elaborar normas nacionales⁹.

Por otro lado, a nivel internacional, se debería difundir la lista de archivos responsables de los puntos de acceso, con la posibilidad de crear grupos de trabajos de archivos de diferentes países.

Debe tenerse en cuenta que este Código se refiere al archivo responsable de la creación del encabezamiento y no al código de referencia identificativo del archivo que conserva la documentación al que se refiere la Norma ISAD (G). Quizás sería oportuno denominar este código de la Norma ISAAR (CPF) de otra manera, como código de control o código del responsable, con el fin de no utilizar el término identificación que figura en el área primera (Área de Identificación) de la Norma ISAD (G).

En cuanto a la *entrada autorizada*, está planteada según las reglas de catalogación bibliotecaria nacional, así pues podemos encontrar algunas diferencias. Por eso, el intercambio internacional en principio no parece que sea totalmente perfecto.

No obstante, sí se ha pensado que la diversidad lingüística no sea un obstáculo al intercambio de información, pues el encabezamiento autorizado puede respetar las diferentes lenguas eligiendo una lengua para la entrada autorizada y el resto utilizando el

campo denominado de entrada(s) *paralela(s)*. También podía considerarse la utilidad de este campo cuando existen varias lenguas en un estado, una entrada podría estar en una lengua referenciando las otras en el otro campo, y viceversa. Sería el caso por ejemplo de España, con la documentación producida primero por el Estado y luego por las comunidades autónomas que tengan su propia lengua.

Tampoco podemos olvidar el caso de los organismos internacionales, cuyas denominaciones a veces cambian según las traducciones. De ahí, la importancia en seleccionar la mejor traducción en la lengua del archivo responsable de registrar la entrada autorizada. Asimismo, es preciso considerar la utilización de siglas y abreviaturas en las entradas autorizadas y en las paralelas cuando se trate de entidades nacionales o internacionales. En el caso de las siglas homónimas se puede optar en las entradas paralelas por el inglés para las organizaciones de origen anglosajón, por el francés para las de origen francófono.

Los encabezamientos autorizados de personas o de familias, al igual que las entidades, también son conforme a las reglas de catalogación bibliotecaria nacionales (en España, apdo. 5.1), por eso se debe tener en cuenta lo siguiente:

- las variaciones ortográficas y lingüísticas para las entradas paralelas,
- la necesidad de paréntesis aclaratorios si un padre y un hijo se llaman igual por ejemplo, o poner los dos apellidos,
- en el caso de las casas reinantes en más de un país, se debería señalar en la entrada autorizada, si nos referimos o no a un país en concreto, por ejemplo entre paréntesis.

Por otra parte, es muy necesaria la claridad del archivero para considerar cuando se trata de un fondo documental de un individuo o de una familia, pues en ocasiones esa delimitación es difícil en los archivos familiares¹⁰.

Respecto a las *formas excluidas* de la entrada autorizada, deben abarcar todas las posibilidades de denominaciones y las siglas o abreviaturas no utilizadas en la entrada principal, ya sean nacionales o extranjeras si se trata de organismos internacionales.

Las *entradas relacionadas* son muy importantes, ya que las entidades se denominan tal y como fuera su nombre en el momento histórico de la documentación. De este modo, se pueden conocer los diferentes nombres de una misma entidad a lo largo de la historia. Sucede lo mismo con las que puede relacionarse su documentación, tanto si se trata de una entidad, como si nos estamos refiriendo a una persona o una familia, pues, por ejemplo, la correspondencia de los archivos personales se complementa, o bien en el caso de las familias, los distintos enlaces conectan los árboles genealógicos de unas con otras, aunque en los anexos no se expresa ningún ejemplo de estos casos.

Este campo de entrada relacionada en la Norma comprende sólo el *véase también*, es decir, la relación jerárquica a mayor y la jerárquica a menor, y a su vez estas entradas relacionadas serán entradas principales. No podemos olvidar que estas relaciones son fundamentales para el seguimiento de las series documentales y para establecer posteriores guías de fuentes documentales. De hecho, en la norma ISAD(G), en su Área de Documentación Asociada (3.5) se prevén las unidades de descripción relacionadas en el archivo (3.5.3) y la documentación complementaria en otros archivos (3.5.4).

Por otra parte, se debería evitar la confusión que figura en el ejemplo francés de los anexos de la norma referida a una entidad en la *entrada autorizada relacionada*, que se encabeza por el país según el modo bibliotecario pues, en realidad se trata en este campo (1.6) de señalar el código de identificación seguido de esa entrada autorizada relacionada.

Así mismo, este campo de entrada relacionada se completa con otro del *Área de Información*, el denominado de *relaciones*, el 2.1.8, que comprende el *véase además*, es decir, la relación asociativa. Cabría considerar la posibilidad de considerarlo en el Área anterior para abarcar todas las relaciones posibles, máxime si pensamos en la elaboración posterior de un tesaurus. También sería interesante idear la conexión entre estas relaciones de la Norma ISAAR (CPF) con el Área de Documentación Asociada de la Norma ISAD (G).

Por otra parte, este Área de Información es muy detallada y, aunque se pretende un conocimiento muy exhaustivo de un órgano productor, es evidente que la documentación del Antiguo Régimen precisará más redacción en la información que la más moderna, en donde la redacción debería ser sucinta. Así el campo de *Nombre* llevará las distintas denominaciones de una entidad acompañada por sus fechas extremas. Y en el campo referido al *Organigrama*(2.1.7), se debería detallar cada nombre con su fecha referida a su creación, supresión y modificación (y la de publicación en el boletín oficial del país, en el caso de un órgano administrativo).

Los *finés, funciones y actividades* (2.1.6) deberían redactarse lo más sucintamente posible, y más bien debería optarse por conceptos que se tradujeran en descriptores, para la elaboración posterior de un cuadro orgánico-funcional de la documentación. Esos descriptores, a su vez, deberían aplicarse en el Área de Identificación y, en su caso, en el Área de Contenido y estructura de la Norma ISAD (G).

La redacción, lo más breve posible, debería aplicarse sólo para el campo de *otra información significativa* (el 2.1.9).

Por último, en el *Área de Notas*, el archivero al mencionar las fuentes consultadas debería citar tanto las que contuvieran información como las que no.

Así pues, la descripción del encabezamiento autorizado es exhaustiva y permite conocer al órgano productor. Además, la normalización de la denominación del órgano productor, sea una entidad, familia o persona, la aplicaremos en la clasificación orgánica de los fondos documentales del archivo, así como en el Registro General de Entradas de Fondos, Repertorio de Series Documentales y Relaciones de Entrega si el archivo es administrativo.

Debemos señalar a este respecto que desde hace unos años hemos realizado una tarea similar por idénticos motivos:

En 1990, para la identificación de los fondos de la Secretaría General del Movimiento, y con la colaboración de un equipo de licenciados de la Escuela-Taller del Archivo General de la Administración en Alcalá de Henares, estudiamos toda la evolución de su organigrama empleando para el estudio de las series documentales las relaciones, *véase*, *véase además* y *véase también*, y diseñamos una ficha al efecto: el *véase* se aplicaba a los términos excluidos y remitía al elegido; el *véase además*, a la relación entre los órganos productores, y, por tanto, entre las series documentales, y el *véase también*, a los cambios de denominación o asunción de competencias de otro órgano.

La ficha, además, recogía las fechas de creación, modificación y supresión de cada órgano y la dependencia jerárquica, además de las relaciones señaladas. Así pues, se establecían las relaciones jerárquicas y asociativas entre los órganos, además de dar a conocer otros datos sobre su existencia.

Modelo de ficha
Unidad NORMATIVA Creación: Supresión: Desarrollo: DEPENDENCIA RELACIONES O REFERENCIAS

Esa misma ficha, posteriormente, la aplicamos al estudio institucional del Ministerio de Asuntos Exteriores, que después se transformó en un tesoro institucional con la normalización terminológica de cada órgano y con la codificación para una lista interna de una base de datos (la base ARGE) referida a los campos de organismo remitente y organismo productor¹¹. De esta manera se aseguraba la correcta introducción de datos en las bases de datos y la imposibilidad introducir criterios subjetivos en la denominación de cada organismo, desde la adopción de abreviaturas o siglas, al uso de mayúsculas o minúsculas, etc. Además, los órganos productores conocían la terminología fijada y en las transferencias de documentación informatizada desde cada archivo de gestión al Archivo General del Departamento se garantiza que no hubiera posibilidad de error en la denominación. Y a ello se une que el Registro General de Entrada de Fondos informatizado contiene el vocabulario de los órganos productores y remitentes normalizado.

La normalización terminológica de cada órgano remitente y productor de la documentación, pues a veces uno y otro no coinciden, aplicado al Registro de Entrada de Fondos, tanto manual como informatizado, permite ya un control y una descripción de los ingresos de la documentación en una primera fase de organización archivística. Asimismo, en esta fase de identificación de los fondos documentales se debe elaborar un cuadro de clasificación orgánico de la documentación del archivo, de manera que el vocabulario controlado relativo a los órganos productores del Registro de Entrada de Fondos se aplique también en dicho cuadro.

De alguna manera el planteamiento que hemos llevado a cabo para la identificación de los fondos y las series documentales coincide básicamente con las Normas ISAAR (CPF). Es decir, en la imperiosa necesidad de normalizar el vocabulario del órgano productor y conocer sus relaciones.

LA INDIZACIÓN EN LA DESCRIPCIÓN

Pero no acaba aquí la normalización terminológica en las tareas de la descripción archivística, pues el siguiente paso procede del análisis de las propias funciones, competencias y actividades del órgano productor para elaborar el cuadro de clasificación orgánica funcional del archivo, es decir, de las series documentales. Si se normaliza la terminología de cada función, competencia o actividad, cada serie podrá estar presentada por una palabra clave, es decir, por un concepto elegido. Es más, diversos tipos documentales podrían agruparse bajo un concepto que correspondiera a una función para una posterior indización a partir del cuadro de clasificación. De hecho, en el campo 2.1.6 del Área de Información ya podían señalarse como decíamos anteriormente.

Nosotros aplicamos palabras clave en un estudio sobre un fondo del antiguo Ministerio de la Gobernación¹². Posteriormente mediante palabras clave agrupamos en el fondo de mapas y planos del Archivo General del Ministerio de Asuntos Exteriores los mapas y planos relativos a un mismo término geográfico amplio. Posteriormente, esas palabras clave las convertimos en descriptores.

Así pues, mediante palabras clave podemos recuperar información sobre una función, competencia o actividad, o sobre una materia o asunto. La palabra clave es seleccionada por ser representativa, pero no tienen entre sí ninguna relación. El índice a partir de esas palabras clave se caracteriza por su economicidad¹³, pero es preciso tener en cuenta para la recuperación de la información que si se utilizan palabras clave deben evitarse las sinónimas, es decir, palabras con igual significado, y las polisémicas, o sea, las de significado múltiple.

El control del vocabulario extraído del análisis documental en las tareas descriptivas, sin embargo, puede y debe ir más allá que en la selección de palabras clave, ya que el siguiente paso es su conversión en descriptores al establecer entre cada término relaciones jerárquicas, asociativas y semánticas. Las jerárquicas y asociativas proporcionan exhaustividad y pertinencia en la información; las semánticas, eliminan la sinonimia y la polisemia.

La indización mediante los descriptores es la más completa que puede obtenerse, pues comprende los distintos niveles de descripción de los fondos documentales que abarca el lenguaje documental. El diccionario compuesto por ese vocabulario controlado se denomina *tesauro*, que, al menos, se presenta en dos formas: la alfabética de descriptores con sus relaciones, y la sistemática, con los descriptores en sus relaciones semánticas.

La claridad en un tesauro es fundamental para la recuperación de la información, por eso cuanto más se especifiquen las relaciones de cada descriptor y mejor se señalen gráficamente, tanto más útil será para el usuario, pues constituye un puente entre el documento y el investigador.

Se trata pues de un lenguaje especializado, y por tanto, de un tesauro especializado. Cada archivo dispondrá de uno propio, si bien habrá términos que podrán encontrarse en diferentes tesauros. El ideal será establecer tesauros específicos por fondos documentales, que se conserven en distintos archivos.

Por otra parte, como todo tesauro está abierto a incorporaciones de nuevos términos y a modificaciones, máxime si el archivo es administrativo.

Ahora bien, si los descriptores expresan el concepto seleccionado, también hay términos que de por sí son descriptores. Son los denominados identificadores, y cada tesauro contiene los suyos según el tipo de documentación¹⁴. En la documentación de archivos serían identificadores, en primer lugar, los órganos productores, es decir entidades, personas y familias. De manera que los encabezamientos autorizados archivísticos referidos a esos órganos son descriptores de por sí o identificadores. Otros identificadores serían los topónimos.

Asimismo, los descriptores de un archivo pueden ser simples, es decir, de un término, y compuestos, pues algún concepto requerirá varias palabras y modificadores. Es el caso, por ejemplo, en el tesauro del Archivo del Ministerio de Asuntos Exteriores¹⁵, de la denominación de los tratados: ejemplo TRATADO DE GINEBRA SOBRE DESCUBRIMIENTOS CIENTÍFICOS.

En un tesoro, las principales relaciones son las de equivalencia expresadas por el *use* y *usado por* (*use*, *UP*); las jerárquicas, establecidas por el *término genérico* (*TG*) y el *término específico* (*TE*), y las asociativas, por el *término relacionado* (*TR*). También un tesoro incluye lo que se denomina "relaciones que no comparten reciprocidad" que son sobre toda la *nota de alcance* (*NA*) y el *equivalente lingüístico* (*LE*)¹⁶. Además, según el programa informático, la fecha (*DT*) referida cuando se incluyó el descriptor en el tesoro; el otro (*OT*) que permite incluir un código numérico identificativo del término que será válido para la búsqueda; y el párrafo (*PARA*), para indicar en qué campos de una base de datos queremos que se verifique este término cuando se introducen los datos.

Para la elaboración del Tesoro Institucional del Ministerio de Asuntos Exteriores, utilizando el programa informático BPS, aplicamos además el término principal, término genérico, término específico y el término relacionado, la *DT* para las fechas de creación y supresión, la *NA* para aclaraciones y fecha de modificación y para el *OT* para el código numérico de ese término principal. Las facetes o grupos temáticos del tesoro estaban formadas por las grandes áreas de competencia del Ministerio (ej. consulares, protocolo, etc.).

Experiencias en archivos de lenguajes de indización aún son escasas. Salvo las listas de encabezamiento de materias muy utilizadas en los países anglosajones y con fuerte influencia de la biblioteconomía¹⁷, debemos reseñar el *Tesoro de la serie W* que editó en 1989 la Dirección de Archivos de Francia¹⁸, el *Vocabulario de indización y Consulta de la base de datos BADATOR*¹⁹, el ya citado *Tesoro de Archivo del Ministerio de Asuntos Exteriores de España*, y el reciente *Tesoro del Archivo Municipal de Barcelona*²⁰.

Sin olvidar las normas que rigen la elaboración de un tesoro, como las ISO 2788-1974, *Principios rectores para el establecimiento y desarrollo de Thesaurus monolingües*, y la UNISIST -1984 *Directrices para el establecimiento y desarrollo de Thesaurus monolingües*, para elaborar un tesoro de archivo hay que tener en cuenta las características del *documento de archivo*.

El *Tesoro W* se crea a partir de los descriptores referentes al ámbito de la intervención de la Administración y los términos ligados a las competencias²¹. Se presenta de forma temática, con once grupos temáticos o capítulos que alcanzan tres niveles de jerarquía: rúbrica, subrúbrica y descriptor simple; el tesoro alfabético de materias, y dos listas de autoridad, la de las palabras clave o útiles referentes a operaciones de la Administración, acción administrativa o conceptos vagos, y la de tipos documentales. Este tesoro persigue dos objetivos²²: establecer una descripción inicial somera pero fiable, y cubrir los fondos de la serie W.

El *Vocabulario de Indización de la base de datos BADATOR* se presenta en estructura jerárquica, alfabética y arborescente. Cada descriptor va acompañado de un código numérico de acorde con la estructura jerárquica que responde a un cuadro de clasificación por materias. Abarca documentación desde el s. XVI hasta el XIX.

El *Tesoro Archivo* del Ministerio de Asuntos Exteriores español se presenta en estructura jerárquica, alfabética y con un índice permutado. Integra desde cuadros de clasificación de fondos, hasta organigramas, identificadores, y abarca desde el s. XIX hasta la actualidad²³.

En la introducción de ese tesoro se señalan sus características, pues, aunque se sabe que para la elaboración de un tesoro se prefiere el sustantivo, el singular al plural, el término desarrollado a las siglas y acrónimos..., la documentación exige la adaptación del lenguaje de indización a su especificidad. De esta manera, el tesoro incorpora:

- o Gran número de identificadores, relativos a la estructura de la Administración, órganos del estado, organigrama del Ministerio, tipo de documento, denominación histórica de cada país, los términos geopolíticos más utilizados, los espacios políticos históricos, las embajadas, consulados y representaciones españolas en organismos internacionales, los bancos internacionales, y las exposiciones universales e internacionales.
- o Descriptores compuestos, incluso con fecha añadida como signo distintivo. Es el caso de tratados o de reuniones (congresos, conferencias, etc.) diferentes celebradas en el mismo lugar, como por ejemplo:

REUNIÓN DE EXPERTOS LA VALETTA 1979 <i>SN Sobre cooperación en el Mediterráneo</i> <i>RT Conferencia de Belgrado 1978</i> <i>Mediterráneo</i> REUNIÓN DE EXPERTOS LA VALETTA 1991 <i>SN Sobre arreglo pacífico de controversias</i> <i>RT Conferencia de Viena 1986</i> <i>Mantenimiento de la paz</i>

- o Palabras útiles, que responden a la necesidad de combinar un descriptor de aplicación general (CONFERENCIA, REUNIÓN, etc.) con un descriptor del asunto específico de que se trate.
- o Singular y plural para marcar las diferencias entre tipo documental (CREDENCIAL) y asunto de protocolo (CREDENCIALES).
- o Siglas y abreviaturas conocidas. Piensese, además, en las abreviaturas de unidades de la administración como dirección general, subdirección general, etc.
- o Cuadro de clasificación de fondos. Es el caso del Grupo temático de consulares, de protocolo, de servicio exterior prácticamente de la mayoría de las facetas del Tesoro del Ministerio de Asuntos Exteriores.
- o Organigramas, como el caso de las organizaciones internacionales.

Así pues, la propia estructura del tesoro está marcada no sólo por los temas de competencia del Ministerio sino también por su propio organigrama: consulares, protocolo, servicio exterior... Estructura abierta en función de la necesidad de crear nuevas facetas o grupos temáticos o de la relevancia de alguna en particular.

Por otra parte, la nota de alcance se utilizó para:

limitar el alcance de un término:

ACTIVIDAD DIPLOMÁTICA DEL EMBAJADOR DE ESPAÑA

SN Aplicado a las reuniones, conversaciones, etc., del embajador español con otros embajadores acreditados en un país sobre un tercer país o un asunto determinado, según clasificación de los archivos de las embajadas.

definir el término:

NUEVA GUINEA ORIENTAL

SN El nordeste fue alemán desde 1894 a 1914, el sureste fue inglés desde 1888 a 1906. Después, todo fue de Australia hasta 1975.

fechas:

RONDA DILLON
SN 1960-1962

proporcionar información:

COMUNICACIÓN VALIJA CIFRA
SN Se suprime por Orden Circular 3189 de 15 de noviembre de 1993

El término relacionado informa acerca de:

cambios de denominación de países y organizaciones:

CONGO BELGA
SN Desde 1867 a 1960
En 1960 Congo (Leopoldville), y desde 1966 Congo (Kinshasa)
BT Geopolítica
RT Congo

relación de cada organización, congreso, etc. con su tema:

HONLEA
UF Reunión de Jefes de Organismos Nacionales contra las Drogas
BT NNUU
RT Tráfico de Drogas

Por último, cabe señalar que el programa utilizado para este tesoro, el BRS, permite la edición de un índice permutado, que es muy útil para el manejo del tesoro. Se trata de un índice KWIC (Key word in Context), es decir, el descriptor se sitúa en el centro de cada línea y el texto se reestructura a su alrededor²⁴.

Este mismo programa se está utilizando en una base de datos BIMA, referida a documentos audiovisuales, de varios archivos catalanes, cuyo tesoro presenta determinadas características debido a la índole de esos documentos, la manera de consultarlos y las necesidades diversas de los usuarios. Además de su presentación temática, en diez áreas temáticas, y alfabética, presenta unos anexos sobre descriptores geográficos e identificadores, compuestos por entidades, personas, estructuras arquitectónicas y periodos cronológicos. Y se añade un glosario²⁵.

CONCLUSIONES

En definitiva, la normalización implica una terminología normalizada a su vez. La descripción archivística debe tomar del análisis documental aquellas operaciones consistentes en determinar los elementos de la descripción que se considerarán como puntos de acceso para acceder a la información de una manera pertinente y rápida. Los términos de indización se seleccionan desde la primera fase de la descripción, esto es, en la identificación de los fondos documentales, señalando los órganos productores, sin olvidar los órganos remitentes si son diferentes, y las funciones y competencias de cada órgano productor, de manera que al registrar los encabezamientos autorizados

archivísticos, también puedan señalarse los términos de indización para un cuadro orgánico-funcional de los fondos documentales. Por otra parte, unos y otros serán identificadores y descriptores del tesauo.

Cabe considerar que esta normalización terminológica y descriptiva podrá requerir una metodología diferente según se trate de documentación del Antiguo Régimen o contemporánea.

Ahora bien, sea como fuere el tipo de documentación, el trabajo en equipo es inevitable para la indización idónea y su máxima difusión y utilización. La conexión entre la Norma ISAD (G) y la Norma ISAAD (CPF) constituye una de las experiencias archivísticas más interesantes que se plantean y a las que los programas informáticos no son ajenos.

Mi intención es promover desde la Subdirección General de Archivos Estatales el desarrollo de un sistema informático que contenga las exigencias de esas Normas, creando equipos multidisciplinares de archiveros e informáticos para que la Red de Archivos Estatales comparta la utilización de bases de datos y tesauros, y se elaboren cuantas normas nacionales sean necesarias.

NOTAS

- 1 Norma Internacional de Descripción Archivística, preparada por la Comisión ad Hoc de Normas de descripción (ICA/DDS) en 1993, cuya versión al castellano fue editada por el Ministerio de Cultura en 1995 y se debió a Luis Martínez, Rosana de Andrés y Victoria Arias.
- 2 Preparada por la Comisión ad Hoc de Normas de Descripción (ICA/DDS), que en principio se denominó ISAD (AR), Norma Internacional de Descripción Archivística para las Entradas de Autoridades, y que tras el anteproyecto final de la reunión de la Haya en 1994 se convirtió en el texto actual en la reunión de París, del 15 al 20 de noviembre de 1995. Su versión española, editada por el Ministerio de Educación y Cultura, se debe a Asunción de Navascués Benlloch.
- 3 Incorporadas a las normas por la AENOR, es la Norma UNE 50-121-91: Documentación. Métodos para el análisis de los documentos, determinación de su contenido y selección de términos de indización.
- 4 V.F.W. Lancaster, *El control del vocabulario en la recuperación de la información*, Valencia: Univ. De Valencia, 1995, p. 212
- 5 Id., p. 131
- 6 V. Jean-Yves Rousseau, "La normalisation et la informatization des Archives" en *La Normalisation en Archivistique*. . . Melanges Jacques Ducharme, Quebec, 1992, p. 144-145.
- 7 En este sentido, J.R. Cruz Mundet considera que la renovación del lenguaje administrativo es vital en el proceso de modernización de las propias administraciones y eje de la calidad en los servicios (v. "Los archivos administrativos: un reto para la profesión", en TRIA, n° 3 (1996), p. 31)
- 8 Se ha considerado la posibilidad de que la indización se reservara sólo a los documentos cuya conservación fuera cierta o muy probable. Tal es el planteamiento que se hace en la introducción del Tesauo W, editado por la Dirección de los Archivos de Francia. 1989, p. 11.
- 9 Según el Sistema Informático Estatal, cada archivo dependiente de la Subdirección General de Archivos Estatales tiene asignado un código: 001 Archivo Histórico Nacional, 002 Archivo General de Indias, 003 Archivo General de Simancas, 004 Archivo de la Corona de Aragón, 005 Archivo General de la Administración, 006 Archivo de la Real Chancillería de Valladolid, 008 Sección Guerra Civil del Archivo Histórico Nacional, 008 Archivo Histórico Provincial de Vizcaya, 101 Sección Nobleza del Archivo Histórico Nacional. Y sucesivamente se añadirán códigos a los archivos del sistema español de archivos.
- 10 Las definiciones de un fondo de individuo y de un fondo de familia en Denis Garon, "Le fonds d'Archivers: objet premier de la normalization en Archivistique", en *Normalisation en Archivistique*. . . Melanges Jacques Ducharme, Quebec, 1992, p. 109-110.

- 11 V. Elisa Carolina de Santos Canalejo, *Guía del Archivo del Ministerio de Asuntos Exteriores* (en prensa).
- 12 V. Elisa Carolina de Santos Canalejo con Guadalupe Moreno López, M^a Nueva Yuste Blázquez y Juan Carlos Moreno Duque, *La Administración Local Española 1940-1978: Fondo del Archivo del Ministerio del Interior*, Madrid: Ministerio del Interior, 1990.
- 13 V. Margarita Badillo Nieto y Severiano Aznar Peñarroya, "Técnicas de documentación para la Administración Pública", Madrid, INAP, 1986, p. 93.
- 14 V. F.W Lancaster, ob. Cit., p. 141.
- 15 V. Elisa Carolina de Santos Canalejo con la colaboración del grupo de trabajo del Archivo General y del grupo de trabajo de los archivos de gestión del Ministerio de Asuntos Exteriores, *Tesaurus de Archivo*, Madrid, Ministerio de Asuntos Exteriores, 1994.
- 16 En las normas UNE figuran las siguientes relaciones: nota de alcance (NA), use, usado por (UP), término cabecera (TC), término genérico (TG), término tope genérico (TTG), término genérico partitivo (TGP), término específico (TE), término específico genérico (TEG), término específico partitivo (TEP) y término relacionado (TR).
- 17 V. J.R. Cruz Mundet, *Manual de Archivística*, Madrid, Fundación Germán Sánchez Ruiperez, 1994, p. 296.
- 18 Thesaurus W. *Vocabulaires normalisés pour la description de l'indexation des archives administratives locales temporaires*. Paris, 1989.
- 19 R. Martín, coord.... Bergara; Gobierno Vasco, 1993, 245p.
- 20 Silvia Domenech i Fernández, *Tesaur BIMA: Arxiu Municipal de Barcelona*, Ajuntament de Barcelona, 1997, 127 p.
- 21 Ob. Cit. P. 6.
- 22 Ob. Cit. P. 11.
- 23 Sobre sus características vease la introducción del Tesaurus.
- 24 Otro tipo de índice permutado es el KWOC (*Key word out of context*), en donde el descriptor se sangra del texto.
- 25 Ob. Cit. P. 165 y 55.

BIBLIOGRAFÍA

- AENOR, Documentación: tomo 2, *Normas fundamentales: Recopilación de Normas UNE*. Aenor, 1994.
- AMAT NOGUERA, Nuria, *Documentación científica u nuevas tecnologías de la Información*. Madrid: Pirámide, 1987, 527 p.
- ABIZANDA SANZ, Carolina y PRADO MARTÍNEZ, Miguel Angel del, *La aplicación de los lenguajes documentales en Archivos*, en *Actas de las V Jornadas de Archivos Aragoneses*. Zaragoza, 1994, p. 287-291
- BADILLO NIETO, Margarita, y AZNAR PEÑARROYAS, Severiano, *Técnicas de Documentación para la Administración pública*. Madrid: INAP, 1986, 207 p.
- BONAL ZAZO, José Luis, y ORTEGO DE LORENZO CACÉRES, Pilar, *Norma ISAAR(CPF): cambios en la versión definitiva*, en *Boletín ACAL*, n^o22-23, (1996-1997), p. 20.
- Id., *ISAAR(CPF): Norma Internacional Archivística para las Entradas de Autoridad de organismos, personas y familias*, en *Boletín ACAL*, n^o19, p. 2-6
- BORKO, Harold, Y BERNIER Charles L., *Indexing concepts and methods*. London: Academic Press, 1978.
- CHAUMIER, J., *Análisis y lenguajes documentales: el tratamiento lingüístico en la información documentaria*. Barcelona: Mitre, 1986.
- COOK, Michael, *Problems arising from the application of informatics to archives en Archival Science of the threshold of the year 2000*. University Macerata, 1990, p. 275-295.
- CRUZ MUNDET, José Ramón, *Manual de Archivística*. Madrid: Fundación Germán Sanchez Ruipérez, 1994, 408 p.
- Id., *Los archivos administrativos: un reto para la profesión*, en *TRIA*, n^o3 (1996).
- CURRAS, E., *Thesaurus: lenguajes terminológicos*. Madrid: Paranino, 1991, 284 p.
- DOMENECHI FERNÁNDEZ, Silvia, *Tesaur BIMA: Arxiu Municipal de Barcelona*. Ajuntament de Barcelona, 1997, 127 p.
- GARCIA GUTIÉRREZ, Antonio Luis, *Lingüística documental. Aplicación a la documentación de la comunicación social*. Barcelona: Mitre, 1984.
- GIL URDICIAIN, Blanca, *Manual de lenguajes documentales*. Madrid: Noesis, 1996, 269 p.

- HEREDIA HERRERA, Antonia, *Thesaurus indiano, en IX Congreso Internacional de Historia de América*. Sevilla, 1992, p. 255-260.
- ISO 2788-1974. *Documentation. Guidelines for the establishment and development of monolingual thesauri*.
- ISO 5964-1985. *Documentation. Principes directeurs pour l'établissement et le développement de thesaurus multilingües*.
- ISO 5963-1985. *Méthodes pour l'analyse des documents, la détermination de leur contenu et la selection des termes d'indexation*.
- LANCASTER, Frederick W, *El control del Vocabulario en la recuperación de la información*. Valencia: Universidad de Valencia, 286 p.
- LAUREILHE, M. T. *Le thesaurus : son rôle, sa structure, son élaboration*. Villebaurne: Presses de l'ENSB, 1981, 88 p.
- MARTIN SUQUÍA, Ramón, "El control del lenguaje en el proceso de formación de una base de datos de archivo: IRARGI 1", en *DOCUMAT*; 90, p. 940.
- NORMALISATION en Archivistique, : *Un pas de plus dans l'évolution d'une disciplina: Melanges Jacques-Duchenne, sou la direction de Carol Contrera, avec la collaboration de Frederick Broch, Quebec: Documenter; Association des Archivistes du Quebec, 1992*.
- ROY, Richard, *Introduction a l'indexation Documentaire*. Le Mans, 1985.
- SANTOS CANALEJO, Elisa, con la colaboración del grupo de trabajo del Archivo General y del grupo de trabajo de los archivos de gestión del Ministerio de Asuntos Exteriores, *Tesaurus Archivo*. Ministerio de Asuntos Exteriores, 1994, 474 p.
- Id. el grupo de trabajo formado por Guadalupe Moreno López, María de la Nueva Yuste Blazquez, Juan Carlos Moreno Duque, *La Administración local española 1940-1978: Fondo del Archivo del Ministerio del Interior*. Madrid: Ministerio del Interior, 1990, 209 p.
- Id. *Repertorio Normativo del Movimiento Nacional, AGA, 1990*.
- Id. *Guía del Archivo General del Ministerio de Asuntos Exteriores* (en prensa).
- SOERGEL, D, *Indexing languages and thesauri: Construction and maintenance*, Melville, 1974.
- THESAURUS: *vocabulaires normalisés pour la description et l'indexation des archives administratives locales contemporaines*. Paris, 1989.
- UNISIST-UNESCO, *Guidelines for the establishment and development of Multilingual Thesauri*. Paris, 1976.
- VAN SLYPE, M.G., *Definition des caracteristiques essentielles del thesaurus*, Burian Marcel Van Rijk, 1976.
- Id., *Lenguajes de indización: concepción, construcción y utilización en los sistemas documentales*. Salamanca (etc): Fundación Germán Sánchez Ruiperez, 1991.
- VICKERY, B.C., *Thesaurus: a new word in documentation*, Document, 16, (1968), p. 181-189.
- ZUMALAE, Irune, *Normalización Internacional para el Registro de Autoridades en Archivos para Entidades, Personas o Familias*, en *ALDEE*, junio 1996, p. 25-35.