



**HAL**  
open science

## **Control conceptual en estructuras jerárquicas y combinatorias: CDU v. Bilindex**

Enrique Wulff Barreiro

### ► **To cite this version:**

Enrique Wulff Barreiro. Control conceptual en estructuras jerárquicas y combinatorias: CDU v. Bilindex. Actas de las VI Jornadas Bibliotecarias de Andalucía: Alcalá de Guadaira (Sevilla), 2, 3, 4 de noviembre de 1989, pp.289-298, 1990, 84-404-8528-X. <hal-01170831>

**HAL Id: hal-01170831**

**<https://hal.science/hal-01170831v1>**

Submitted on 6 Jul 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Control conceptual en estructuras jerárquicas y combinatorias: CDU v. Bilindex

**Enrique Wulff Barreiro**

Diplomado en Biblioteconomía y Documentación  
Universidad de Granada

Con la saludable intención de construir un sistema de indización susceptible de mecanización, el Dr. Mortimer Taube y Asociados avanzaron en 1957 una propuesta<sup>1</sup> de elaboración de un diccionario de asociaciones de gran valor tanto para la búsqueda de literatura como en lo relativo a la investigación de operaciones. Estos autores opinaban que el resultado apuntado por un trabajador en investigación de operaciones, durante la II Guerra Mundial, de que aún cuando no abatieran los aviones enemigos, las baterías antiaéreas seguían siendo valiosas dado que hacían disminuir la precisión de los aviones enemigos lo suficiente como para reducir la probabilidad de que echaran a pique navíos mercantes, también se obtendría por medio de un diccionario de asociaciones mecanizado, en el que "la información relevante se presentara adecuadamente indexada o asociada", de modo tal que la asociación de precisión del bombardeo con armamento de los navíos produjera un resultado automático.

En nuestros días el SMART, un proyecto de investigación de sistemas basado en ordenador que utiliza un procesado automático del texto,<sup>2</sup> confirma lo bien sentado de los postulados sobre los que se construyen los diccionarios de carácter asociativo.

El supuesto de máquina asociativa de Taube y Asociados que resolviera el problema antes planteado trabajaría a la manera de los índices de

---

1. TAUBE, M. and Associates, *Studies in coordinate indexing*, vols. I-V, Documentation Inc., Washington D.C., 1953-1957.

2. SALTON, Gerard, *Automatic information organization and retrieval*, New York, McGraw-Hill, 1968.

aparición simultánea, i.e., una lista del conjunto de pares de entrada que pertenezcan al mismo conjunto de entradas de un documento dado. Cabe comprobar que la relación de aparición simultánea entre entradas está en la lógica de las relaciones de equivalencia y similitud, presentando un conjunto de definiciones y teoremas de amplia aplicación en este caso. Más allá del tratamiento dado a esta relación de aparición simultánea entre entradas por el propio Taube, ¿cuál es el valor de lo obtenido mediante su empleo? Sabiendo que 'a' y 'b' aparecen simultáneamente, i.e., que existe un documento X a cuyo conjunto de entradas ambos pertenecen, y que 'b' y 'c' aparecen simultáneamente, i.e., que existe un documento Y (no necesariamente diferente a X) a cuyo conjunto de entradas pertenecen. ¿Qué sabemos de la conexión entre las "ideas" a y c? Nada definitivo; ni siquiera si existe alguna conexión de importancia. Cuando uno está interesado en los documentos cuyo conjunto de entradas contiene tanto 'a' como 'b', ¿ha de leer asimismo los documentos cuyos conjuntos de entradas contengan 'a' y 'c' pero no 'b'? Es posible. ¿Serían los términos directores suministrados por una lista de entradas de aparición simultánea una herramienta eficaz para los trabajadores de investigación o para los que efectúan búsquedas de literatura? Casi con seguridad no. Estas relaciones que ponen en juego propiedades algebraicas elaboradas, sólo cabe solventarlas por medio de dispositivos que tiendan a consolidar los aspectos prácticos de una aproximación eficaz a la cuestión fundamental del tratamiento de temas afines entre documentos.

### Solidez del sistema de indización y sistema de referencias

La cuestión de cifrar la relación entre documentos en base a la afinidad temática viene a plantear la solidez del sistema de indización en términos de evaluación del sistema de referencias. Esto equivale a preguntarse cómo cambiar la asignación de términos temáticos a una búsqueda o "cómo producir perturbaciones en la selección inicial de términos de búsqueda", en palabras de Estrin.<sup>3</sup> Así pues, prescindiendo de

3. ESTRIN, G.: "Maze structure and information retrieval" En: *Proceedings of the International Conference on Scientific Information*, Washington, D.C., Nov. 16-21, 1958.

planteamientos que incluyan la formulación y reformulación de la pregunta de búsqueda, lo importante es que el sistema venga acompañado de un dispositivo que mantenga de forma estable las asociaciones entre contenidos temáticos pertinentes. Un sistema de indización puede muy bien mantenerse o echarse a perder en función de la calidad de sus referencias. Un dispositivo de este orden tiene, como es obvio, la función que consiste en llamar la atención sobre los sinónimos, pero además, lo que hace, o para lo que en principio está pensado, es ofrecer al usuario la posibilidad de cambiar la formulación de su pregunta de búsqueda a una más o menos estrechamente relacionada y esto con la intención de obtener respuesta relevante no sólo al campo temático producto del cambio sino también al original, i.e., al de partida. En este sentido es importante tener en cuenta la opción de insertar mecanismos de articulación de materia como las que utiliza el sistema Lynch.<sup>4</sup> El presupuesto de partida, en todo caso, es que no hay razones serias para suponer que documentos con distancia  $n$  deban ser necesariamente más relevantes que los documentos con distancia  $m$ , donde  $n$  es más pequeño que  $m$ , ni siquiera que resulte probable que algo semejante suceda de este modo. Se plantea en este contexto, la Teoría de proximidad, el problema de la distancia, de la decisiva importancia de definir una medida de distancia entre campos temáticos adecuada.

### Teoría de proximidad, relación entre términos e indización temática

Fairthorne<sup>5</sup> avanza que "la distancia, en todo contexto, se basa en la idea de 'diferencia'". Mooers<sup>6</sup> se ocupa más de medidas de distancia entre documentos que entre campos temáticos. Vickery<sup>7</sup> opina que "la relación entre términos es el problema semántico central de la indización temáti-

4. Haciendo referencia a la Indización articulada de materia (Articulated subject indexing).

5. FAIRTHORNE, R.A.: "Delegation of classification" En: *American documentation*, vol. 9 (1958), pp. 159-164.

6. MOOERS, C.N.: *Retrieval by the method of proximity transformation*, Cambridge, Mass., 1958.

7. VICKERY, B.C.: "The structure of information retrieval systems" En: *Proceedings of the International Conference on Scientific Information*, Washington, D.C., Nov. 16-21, 1958.

ca", pero no alcanza a definir una métrica de la conectividad. En opinión de Bar-Hillel<sup>8</sup> el fracaso de estas tentativas se debe, no a la falta de envergadura de los procedimientos, sino a que el propio intento es (casi) imposible, desde esta óptica incluso el objetivo de la ICSI<sup>9</sup> de establecer un concepto de "la conectividad de los documentos en una colección junto con una métrica adecuada para definir el grado de relación" es insostenible.

Para solventar estas dificultades ha de recurrirse a la estimación de la probabilidad de que un documento contenga material pertinente al campo temático requerido, a partir de la caracterización del documento por medio de un conjunto de entradas y la caracterización del campo temático requerido por medio de un conjunto de términos controlados, esto es, al concepto clasificadorio de relevancia.<sup>10</sup> Hablando en términos generales, el tiempo destinado a la lectura de un documento vale la pena si y sólo si su grado de relevancia es superior a un cierto punto; de nuevo a grandes rasgos, la estimación del beneficio que se obtiene del examen detenido de un documento ha de ser superior a la utilidad del esfuerzo invertido en su lectura. El propósito de determinar las medidas de relevancia se concreta en la asignación de ponderadores<sup>11</sup> a las entradas pertinentes a la descripción del contenido temático de un documento dado. La operación de ponderar determina la comprensión de la 'medida' en términos de creación de significado, al modo en que lo propone Small.<sup>12</sup> De forma tal que el grado en que un término concreto de un

8. BAR-HILLEL, Y.: *Language and information: selected essays on their theory and application*, Reading, Mass., Addison-Wesley, 1973.

9. *Proceedings of the International Conference on Scientific Information*, Washington, D.C., Nov. 16-21, 1958. National Academy of Sciences — National Research Council, Washington, D.C., 1959.

10. La obra de Gernot Wersig, *Terminology of documentation*, define 'relevancia' ("La calidad de los documentos recobrados mediante un sistema de recuperación para satisfacer las necesidades de los usuarios").

11. La misma obra define 'ponderador' ("Un cuantificador que expresa el significado que tiene un descriptor en relación con el término de indización completo").

12. SMALL, H.G.: "Cited documents as Concept Symbols" En: *Social Studies of Science*, nº 3 (1987), p. 327-340.

vocabulario controlado determinado abarca al concepto tratado en el documento es comprensible desde el estudio del significado y de la referencia de los signos.

### La operación de indización de materia respecto de la de recuperación de la información

Siguiendo a Serrai,<sup>13</sup> el acto del indizador implica una reducción — todo menos obvia— del significado total de documento y su traducción en una fórmula lingüística destinada a servir de trámite para el que busca el documento. Como señala Ullmann, la causa más frecuente de la restricción es la especialización del significado en un entorno concreto, lo que las más de las veces da meramente lugar a la polisemia, dificultad que incide sobre la operación de recuperación de la información. En efecto, la elección de la materia se desarrolla en base a interpretación y transformación conceptual de tipo intensional, esto, en la práctica, se traduce en una fórmula referida a un individuo o clase, y por ende de tipo extensional; quien partiendo del significado interroga el catálogo, recorre un camino inverso. Claramente la operación de indización tiene que ver con el concepto cualitativo cointensional.

### Definiciones

Definimos la Clasificación Decimal Universal, como un sistema de catalogación por materias en el que la combinación de los distintos términos que forman el encabezamiento se hace en el momento de asignar el signo al documento, representando este signo una noción genérica que se despliega en otras más específicas, de carácter enciclopédico. Se entiende por lista de encabezamiento de materia aquel lenguaje asimismo precoordinado de estructura combinatoria es decir en el que se utilizan palabras del lenguaje natural, pero dotadas de los medios de control necesarios para evitar los equívocos y las ambigüedades. En base al primer lenguaje se construye el catálogo sistemático, en base al segundo el alfabético de materias.

13. SERRAI, A.: *Del Catalogo alfabetico per soggetti : semantica del rapporto indicale*, Roma, Bulzoni, 1979.

## Ventajas y desventajas

Las ventajas que presenta el catálogo alfabético de materias, se refieren a la facilidad de consulta y a la posibilidad de reunir los diversos aspectos de una materia; por contra, el catálogo sistemático ofrece la solidez de una designación que supera la fragmentariedad de la materia. Siguiendo a Hunter,<sup>14</sup> el catálogo sistemático se ordena, en general, según el esquema utilizado para la ordenación por estanterías en la biblioteca, siendo desde luego posible utilizar una clasificación genérica en las estanterías, que permita su fácil mantenimiento, y una más detallada en el catálogo para la recuperación rápida de items específicos. Sin embargo, el catálogo alfabético de materias puede complementar a la ordenación sistemática por estanterías reuniendo (mediante el uso de 'relaciones distribuidas') facetas de una materia que el sistema de clasificación ha tenido que dispersar. El acceso a la materia que ambos sistemas proporcionan viene calificado en el caso del catálogo alfabético de materias de "directo" y en el del sistemático de "indirecto", dado que aquí siempre habrá al menos dos pasos; ha de indicarse, no obstante, que Butcher<sup>15</sup> encuentra que el acceso no es directo en el primer caso en un 30% de las ocasiones investigadas.

## Algunos datos de uso

Los resultados del ASLIB Cranfield Research Project, realizado en el Instituto de Tecnología de Cranfield para investigar la eficiencia comparativa de cuatro sistemas de indización de materia (un Catálogo alfabético de materia que usa un registro autorizado de encabezamientos de materia, Catálogo sistemático que utiliza la Clasificación Decimal Universal, una clasificación por facetas especialmente compilada y un sistema postcoordinado), fueron que ninguno de los sistemas eran significativamente superior a los otros y que aunque el control de sinónimos y de la mezcla de los finales de las palabras mejoró el rendimiento de los

14. HUNTER, E.J.: *Cataloguing : a guidebook*, London, Bingley, 1983.

15. BUTCHER, S.J.: "Cataloguing in municipal libraries", En: *Cataloguing principles and practice : an inquiry*, London, Library Association, 1954, p. 93-103.

términos sencillos del lenguaje natural, todo otro método de control de vocabulario no lo hizo. El Estudio Comparativo del Catálogo de la Universidad de Bath (BUCCS) aporta el dato que contempla la superioridad del catálogo sistemático sobre los catálogos elaborados por ordenador, en el sentido de que su tasa de acierto<sup>16</sup> no era más baja.

La encuesta sobre Sistemas de catalogación e indización de materia utilizados en 334 bibliotecas de Gran Bretaña e Irlanda llevada a cabo por el personal de la Escuela de Biblioteconomía y Estudios de la Información del Politécnico de Liverpool en 1976/77 indica que en bibliotecas académicas se utilizaban 122 catálogos sistemáticos frente a 16 alfabéticos de materia, en bibliotecas públicas 123 frente a 14, en especializadas 25 frente a 16, y en la Nacional 0 frente a 3. Por otra parte, y en los USA, dos encuestas proporcionan bastante información de interés relativa al uso de los encabezamientos de materia de la Biblioteca del Congreso (LCSH).

Kanwischer<sup>17</sup> interrogó a 108 bibliotecas universitarias, el 78,2% respondió que consideraba adecuados los LCSH. Enyingi<sup>18</sup> encuestó 200 bibliotecas jurídicas de los USA y encontró que el 85,5% utilizaban los LCSH.

Por último, cabe destacar el que entre los resultados Cranfield aparezca el hecho de que los diccionarios de sinónimos mejoren el rendimiento pero que otros tipos de diccionarios, como los jerárquicos, sean menos eficientes de lo esperado; aspecto confirmado por los experimentos SMART, con lo que de primacía de lo combinatorio sobre los jerárquicos supone.

16. La obra de Gernot Wersig, *Terminology of documentation*, pone en relación los conceptos de Tasa de acierto y de Factor de utilidad.

17. KANWISCHER, D.: "Subject headings trauma : making do with first aid" En: *Wilson library bulletin*, 49, 9 (May 1975), p. 651-654.

18. ENYINGI, P.: "Subject cataloging practices in American law libraries" En: *Law library journal*, 68, 1 (February, 1975), p. 11-17.

### Definición de especificidad documental

Entendemos por especificidad documental, siguiendo a Wersig,<sup>19</sup> la proporción entre el número de documentos no pertinentes y no recuperados y el número total de documentos no pertinentes del registro.

### Especificidad documental comparada: CDU v. Bilindex

La capacidad de la Clasificación Decimal Universal como sistema de recuperación no necesita ser demostrada en absoluto, cabe sin embargo argumentar sus posibilidades frente a una lista de encabezamientos como la edición bilingüe de los LCSH el Bilindex.<sup>20</sup> Centramos esta discusión en las desventajas posibles que la favorable especificidad documental observada en los encabezamientos del Bilindex puede tener cuando se usa respecto de un catálogo sistemático. Las dificultades de control conceptual se presentan cuando se hace necesario consultar varias series de descriptores bajo una sola notación numérica antes de que se encuentre la materia específica que se busca. Esta cuestión apunta al análisis de lo que venimos exponiendo a lo largo de esta contribución, los diversos niveles en que actúa la indización temática; permite dar cuenta de una importante distinción sin cuyo dominio es difícil no incurrir en confusiones y equívocos: la diferencia entre sujeto y objeto del documento, entre el discurso que forma el contenido de una obra, y el tema o cuestión sobre el que dicho discurso versa. Esta forma de entender la materia de un libro, propia de los sistemas precoordinados de estructura combinatoria, puede observarse con detalle discutiendo el siguiente ejemplo: "Los moriscos y la medicina : un capítulo de la medicina y la ciencia marginadas en la España del siglo XVI / Luis García Ballester ; documentos por Rosa Blasco y Luis García Ballester. — Barcelona : Labor, 1984". Cuando se analiza aquí la materia de la monografía contemplándola desde el punto de vista de la historia de la ciencia, al asignar el encabezamiento 'Medicina árabe-Historia-España-Siglo XVI', teniendo en cuenta que la obra contiene

reproducciones de documentos de la época que tratan de medicina y que por tanto el discurso médico como disciplina no está ausente del material a indizar, surge la posibilidad de tomar como asunto u objeto de discurso lo que respecto del mensaje que formaba el contenido 'Medicina árabe' deviene ahora material discursivo desde la perspectiva de la historia de la medicina morisca. La mencionada distinción entre significado y denotación, entre la información que constituye el contenido de la obra y la materia cuya expresión sintética se trata de facilitar mediante el uso de la lista de encabezamientos implica pues una operación, por encima de la habilidad catalogadora, sin cuya comprensión global no resulta factible dotar de capacidad comunicadora al acceso alfabético de materia. El asunto que trata esta obra interesa, por otra parte, al campo designado por la entrada 'Moriscos' cuyo control semántico asegura una significación bien delimitada y constante; la pertinencia de este término, el lugar que ocupa en posición directora dentro del encabezamiento, permite abarcar con más eficiencia el concepto tratado en el documento aumentando por ello el grado de indización del mismo. La cuestión que tratamos, la distancia entre el discurso de un documento y el tema sobre el que versa dicho discurso, puede abordarse a sí mismo desde la óptica jerárquica propia a la clasificación. La solidez que aporta el catálogo sistemático en cuanto a sus entradas tienden a superar el carácter fragmentario de la materia, encuentra necesario recurrir, dada la capacidad de control conceptual pero no terminológico de la Clasificación Decimal Universal utilizada, a establecer relaciones asociativas aseguradas en la sintaxis preestablecida por los dos puntos(:)<sup>21</sup> Ahora bien, la naturaleza precoordinada y jerárquica del lenguaje impone, para que la asociación —necesariamente de carácter rígido— que la CDU permite sea coherente, una relación temática directa so pena de alterar la sintaxis. En este sentido se procede a utilizar el auxiliar común de forma (091) —Historia de una ciencia— con el Índice 61 —Medicina—, permitiendo de este modo relacionar la notación 61(460)"15"(091)(=927) con 946.0"15"(=927)

19. WERSIG, G.: *Terminology of documentation*, Paris, Unesco, 1976.

20. *Bilindex. A bilingual Spanish-English subject heading list. Spanish equivalents to Library of Congress Subject Headings*, Oakland, California Spanish Language Data Base, 1984.

21. TURUGUET, D.: "CDU frente a tesoro en la indización temática para la automatización de una biblioteca científico-técnica" En: *Segundas Jornadas Españolas de Documentación Automatizada*, Torremolinos, Nov. 20-22, 1986, p. 275-285.

para formar la entrada bidireccional 61(460)“15”(091)(=927):946.0“15” (=927). El auxiliar de forma aplicado al índice 61, al operar como dispositivo que marca la distinción entre tema y discurso del documento (capacidad garantizada en el índice de la CDU que, al desempeñar tareas de índice de relaciones, lista el aspecto histórico de la materia bajo su entrada principal), permite dotar de significado coherente a la relación asociativa que se establece para formar la entrada en la secuencia sistemática.

#### **Otras líneas de investigación referidas a los aspectos expuestos en la presente contribución**

Las emprendidas por el Centro de Investigación sobre el Catálogo de la Universidad de Bath que comprenden, contenido y fuentes de los datos catalográficos para uso local; impacto de la Catalogación recomendada; y suministro de acceso de materia a los catálogos de bibliotecas a través de índices de palabras clave.

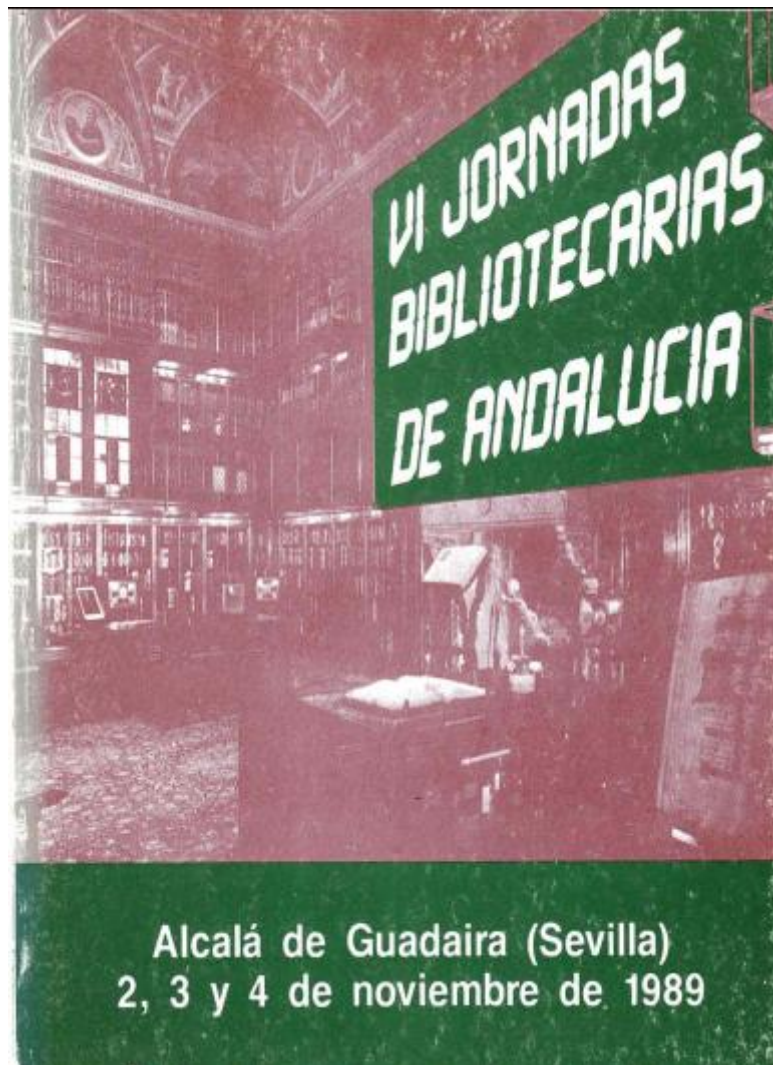
Las derivadas del mantenimiento del catálogo referentes al reemplazamiento de encabezamientos de materia anticuados y a la creación de nuevos subencabezamientos.

Destacar finalmente, como lo ha hecho Arthur Maltby,<sup>22</sup> los inconvenientes que respecto de la investigación sobre el catálogo supone la ausencia de normalización en cuestiones como el código que se utilice, el plan de clasificación que se use, las reglas de intercalación, la forma y tipo de catálogo, y la disponibilidad de un índice de materia o de otro instrumento para el catálogo sistemático.

---

22. MALTBY, A.: "Measuring catalogue utility" En: *Journal of librarianship*, 3-3 (July 1971), p. 180-189.

## **Conclusiones de las Mesas Redondas**



JORNADAS BIBLIOTECARIAS DE ANDALUCÍA (6ª. 1989.  
Alcalá de Guadaira)

Actas de las VI Jornadas Bibliotecarias de Andalucía :  
Alcalá de Guadaira, 2, 3, 4 de Noviembre de 1989. —  
Málaga : Asociación Andaluza de Bibliotecarios, 1990

D.L. GR-1904-1990. — ISBN 84-404-8528-X

1. Biblioteconomía - Congresos, conferencias, etc.  
I. Asociación Andaluza de Bibliotecarios  
02(063)

*Edición de las Actas:*

J. F. Herranz  
Antonio Martín Oñate  
Rafael Olivares

Esta edición se ha realizado con la colaboración de la  
Universidad de Sevilla

© Asociación Andaluza de Bibliotecarios  
Edita: Asociación Andaluza de Bibliotecarios  
C/ Granados, 10, 3º D - 29 00 8 - Málaga  
Tef.: (952) 21 31 88 - FAX 21 31 88

ISBN 84-404-8528-X

D.L. GR-1904-1990

Fotocomposición: Lozano  
Santo Tomás de Villanueva, 14, 4º 3º Granada

Imprime: Proyecto Sur de Ediciones, S.A.L.  
San Juan, 2 (Camino Bajo)  
18100 Armilla (Granada)