Spanish Research in Knowledge Organization (1992-2001)‡

María José López-Huertas*, Evaristo Jiménez Contreras**

Facultad de Biblioteconomía y Documentación, Universidad de Granada (Spain),
*E-mail: mjlopez@ugr.es, **E-mail: evaristo@ugr.es

Dra. López-Huertas is president of the ISKO Spanish Chapter since 1998 and professor at the Faculty of Library and Information Science of the University of Granada (Spain). She teaches knowledge organization, indexing languages and classification. Her research interests are related to the design and construction of knowledge structures for information retrieval and terminology and knowledge organization in interdisciplinary subjects.

Dr. Jiménez-Contreras is former director of the Library and Information Science department at the University of Granada (Spain). At the present he is professor at the same university and manager of the "Assessment of Science and Scholarly Communication" research group. His main interest in Science is Bibliometrics/Scientometrics and Science assessment. In recent years he has written some papers on this subject published in *Scientometrics*, *Research Evaluation* or *Journal of Medical Library Association*, among others.

ABSTRACT: This study gives an initial appraisal of research activity in Spain surrounding "knowledge organization." The sample comprises articles, monographs and PhD. dissertations identified in the following data bases: ISI, LISA, ISOC, REBIUN, RUECA and TESEO. In Spain, "knowledge organization" is a consolidated subject area that shows increasing productivity, although it cannot be considered well developed by any means. A small number of highly productive authors are responsible for the bulk of output. Most research activity stems from university departments and schools of library science, in particular the Universities of Madrid and Zaragoza. A general interest in the theoretical aspects of classification systems, documentary languages and thesauri can be seen.

I. Introduction

Research activity in Library and Information Science (LIS) in Spain can be said to have reached a state of maturity, after some decades of university studies in this area. There has also been qualitatively and quantitatively important scientific output, as shown by a number of studies (Moya & Jiménez Contreras, 1998; Delgado López-Cózar, 2000, 2002; Jiménez Contreras, 2003). Limited attention has been given to the situation of the research carried out in Spain specifically regarding Knowledge Organization, the most recent study being in 1998 (Sánchez Casabón

& García Marco, 1995; Frías & Romero, 1998). For this reason, we believe it necessary to look into this subject area with an updated and more comprehensive analysis.

1.1. Justification and Objectives

There are three reasons why we focus on the period January 1992 to December 2001:

– A ten-year period is long enough to allow us to observe the evolution of the subject of study.
– This is considered to be a period of maturity in this discipline. Other authors also argue that li-

brary and information science emerges as a scientific domain in Spain around the year 1990 (Pérez Álvarez-Osorio, 1997).
– A previous study of similar characteristics covered the analysis of contents and documentary languages from 1982 to 1994 (Sánchez Casabón & García Marco, 1995), and will serve as a case for the comparison of results.

"Knowledge organization," on the other hand, is a field whose boundaries are not always clearly defined. It stands as a broad concept, and not all authors make their understanding explicit. Bliss coined the expression in 1933, in his book the "Organization of Knowledge in Libraries" and described it along semantic lines as a specific area that "studies laws, principals and procedures by which the specialized knowledge of any discipline is structured." It has since been defined in more functional terms as "the representation and organization of documents in information systems" (Andersen, 2002). Either way, its outline is not easy to trace within LIS. It suffers from low visibility in bibliometric studies, as it usually appears subordinated to other categories and may be overlooked when searching with a reference to the contents of research. Hjørland speaks of this problem and traces it to a lack of instruction in specific methods to be used in research on knowledge organization, above all concerning classification (Hjørland, 2002).

In this study, we restrict the realm of KO to those systems that are built on a structure of terms and concepts, such as classifications, thesauri, subject headings, etc. Algorithmic and quantitative systems were not considered in this study. The notion of structure is essential, as we focus on research involving any theoretical, methodological, practical or professional attempt to elaborate structures for the conceptual organization of knowledge, in specialized or encyclopedic settings. Documentary analysis and indexing in a strict sense are therefore excluded from our realm of observation. Moreover, we shall only consider the output of researchers born in Spain or naturalized as Spanish citizens or those working in a Spanish Institution at the time they publish a paper.

We should also discern between research and publication. There is a need to know what is published in a certain topic area without applying overly restrictive criteria a priori, so as to later evaluate which part thereof actually qualifies as scientific research. This is an important distinction, as other authors have pointed out (Delgado López-Cózar, 2002). Further-

more, the comparison between what is published and what is true research will help us appraise the situation overall.

1.2. The state of the art

Studies on LIS research are relatively recent in Spain. Overviews are provided by Jiménez and Moya (1997), Moya and Jiménez (1998, 1999), and Delgado López-Cózar (2000). Meanwhile, Cano (1999) focuses his study on only two sources of information: the Spanish Journal of Scientific Documentation, and the Documentation of Information Sciences. Frías and Romero look at the period 1992-1993 (1998), and finally Jiménez (2002) studied the international output of Spanish authors from 1992 to 2002. A study by Moya (2000) emphasizes the retrieval of information with specific reference to documentary languages, and Sánchez Casabón and García Marco (1995) look at content and documentary language analysis in Spain from 1982 to 1994. There is therefore little overlap in the literature, and the period of study we have chosen clearly requires further attention.

2. Methodology

When assessing scientific output in a special field, the most usual approach is to collect the articles from international databases. However, as our discipline involves other documentary forms such as monographs, conferences, etc., these forms were also included in our search. The documentary types studied, then, are: monographs, dissertations, conference papers (national or international), and articles of any length published in all journals indexed by the databases specified below. It is important to note that publications dealing with KO from an exclusively quantitative perspective and related more directly with information retrieval were excluded from this preliminary analysis, though they will be addressed in a related study of publications dealing with quantitative or algorithmic methods.

2.1 Databases: challenges and limitations

The sources chosen to identify publications were the ISOC (*Centro de Información en Ciencias Sociales y Humanidades del Consejo Superior de Investigaciones Científicas*), LISA (Library and Information Science Abstracts) and ISI for articles and conference papers; TESEO (*Bases de Datos de Tesis Doctorales*) for dissertations; and REBIUN (*Red de Bibliotecas Univer-*

sitarias Españolas) and RUECA (*Red Universitaria Española de Catálogos Absys*), for books.

The first objective was to attain optimal exhaustivity on a national level (ISOC) and on an international level (ISI and LISA). In applying this criterion to books, TESEO was consulted for dissertations, and the collective catalog of the Spanish Universities REBIUN, complemented by RUECA for other monographic works. An ISBN search would have been less reliable, as well as more complicated for reasons of a structural nature, as retrieval by subject is difficult.

In searching the data bases five basic questions were in mind:

– Who researches and publishes?
– How much research is published?
– Where is the research published?
– When was the research published?
– What was the topic of the research and publication?

Some familiar problems arose in the identification of relevant publications:

a) *Poor visibility of KO among bibliometric studies.* The tags are not explicit enough in describing the area, and effectively subordinate it to other subject headings or topical areas. Publications tended to elude the search strategies used.

In the field of LIS in general, KO, classification, thesauri and certain other topics do not appear with a distinctive tag, but are mixed in among the array of categories proposed for the study of research contents. This is well documented by Delgado López-Cózar (2000, 2002) and Cano (1998). At best, KO is included as a subcategory. The authors of bibliometric studies, however, do not consistently subordinate these contents to the same categories, which further increases the difficulty of their identification (Delgado López-Cózar, 2002). For instance, we may find KO under information storage and retrieval, tagged as classification and indexing, leading us to assume that thesauri and KO in general are included. Or we might find it under bibliographic control, which embraces classification, indexing, thesauri and cataloguing, thereby introducing a great deal of noise. More precision can be seen when the scope is limited to IR-related themes (Moya, 2000) including a category for the analysis of documentary languages, or when there is an almost exclusive focus (Sánchez Casabón & García Marco, 1995).

In the opinion of Hjørland, this means that researchers in classification are not as visible on the bibliometric maps proposed for LIS output (for instance by White & McCain); and he believes that this lack of formal recognition is related with the type of research methodology taught at most LIS Schools (Hjørland, 2002). Standard behavioral methodology, qualitative or quantitative, would be more appropriate for a study of users than for a study about classification, because the notion of "users" is compatible with the social sciences. There is a dire need to develop and teach scientific methods geared to the field of research in question.

The lack of visibility is a tendency seen for Spain, but not limited to our territory. Other causes of poor visibility are related with the objective productivity of the authors who conduct research in the area of KO, and the productivity of those investigating other aspects of LIS.

b) *Database design and structure.* There is no doubt that the lack of homogeneity in the design and structure of databases has undesirable effects on information retrieval, especially in the subsequent processing of data for bibliometric purposes. Whereas the structure and the formats used by the ISOC, ISI and LISA include fields such as the institutional origin of the authors, in REBIUN this field does not appear. It is therefore impossible to group authors by origin in the case of books. We acknowledge that our study might suffer from a lack of consistency in this sense.

c) *The heterogeneity of the tools of thematic description in the databases, and the inconsistencies and or imprecisions of indexing.* These problems are familiar and greatly affect sample selection. If the indexing policy of a database is not known, and the indexing language cannot be consulted at the time of the search, the detours and losses of information can be considerable, even when there is no error on the part of the party responsible for indexing itself. It is not easy to find tags that ensure pertinent retrieval. In TESEO, very generic descriptors are used, resulting in extremely noisy searches, requiring many hours of work to cut through the jungle of irrelevant references.

d) *Recall.* It is one of the most common complaints about databases and information systems (López-Huertas, 1997). Recall rates are known to be low, and this hampers bibliometric-related work, inevitably reliant upon databases. Solid familiarity with the databases used will enhance the exhaustivity of a search, in addition to making unnecessary the *a posteriori* distillation of references. For example, with REBIUN we opted to search by "classification"

alone after we observed that other terms like "classi-fication system" or "bibliographic classification" did not retrieve the expected information.

2.2. Obtaining and managing the data

For the selection of documents, thematic searches were done by terms or by classification codes. The databases were queried with a previously established list of terminology, although some formulations and strategies had to be modified.

International output was identified by consulting the ISI database. Two subgroups were defined – that of specialized journals, under the category "Library and Information Science," and another with the rest of the journals and categories. When we refer to the LIS output and aspects such as the origin of the arti-cles or the subjects of research, we refer to the works of the first subset. When we refer to the exportation of our methodologies to other realms, we refer to the remaining journals. To locate the works, two differ-ent strategies were used. In the case of the work pub-lished in LIS journals, we qualified the field "ad-dress=Spain" with all the journals belonging to the category Library and Information Science, according to the lists of the *Journal Citation Reports* (*JCR*) for the last four years. The items located would fulfill the dual condition of being published in LIS journals and being produced by Spanish authors, though not necessarily from a LIS setting. This combination en-tails the drawback that it might ignore journals that were included in the *JCR* before 1997 and later dis-appeared. For the relevant articles published in non-LIS journals, the only possible strategy was a subject category search. The procedure followed in this case was to look for meaningful words that describe our activity and appear in the "BASIC INDEX" of the *Social Science Citation Index* (*SSCI*) and the *Science*

Citation Index (*SCI*). We also adopted the restriction that they be Spanish authors, as in the study by Jiménez (2002). Table 1 shows a list of retrieval terms that is based, in part, on the expressions iden-tified in a study of terminology used for the concept of documentary languages (López-Huertas, 1991).

The LISA query was done using the *a priori* list of Table 1. The results of both searches had then to be filtered, as most of the authors were not Spanish. The search strategy in the ISOC database was based on the classification given by the database in its interface. This interface was later modified and improved, but in this description we limit ourselves to its search po-tential at that time. An *ad hoc* numerical and hierar-chical classification was used, so that the sequences selected related more generally with the topics related to KO, as defined in this study. The codes used in the searches are given in Table 2 below. Because the hier-archical classification was used, the main numerical values are given, along with their generic meaning, and whether searches were complete (that is, if all the classified documents among the maximum and the minimum values indicated were retrieved) or partial. We also distinguish between the documents retrieved and those we finally used in this work.

Monographs were selected from searches in RE-BIUN and RUECA. We discarded the idea of search-ing in the ISBN database because its structure makes retrieval by theme impossible. No distinction was made between authors, publishers and compilers, so that all served as units for computation. Table 3 gives the list used for search by topic.

The result of these searches had to be filtered be-cause these catalogues do not allow the combination of searches by subject with the author's origin as do the journal article databases. There was therefore considerable noise in the retrieval set.

Once selected, the references were exported to Procite for processing. There, they were subjected to an authorship control process to correct for repeated entries and names in the wrong order, problems that are common in ISI and other international databases (Ruíz Pérez, 2002). Normalization presented some additional problems because the first surname of the authors in international databases usually comes after the initials of the first and middle name, yet Spanish authors may be listed under their second surname, we had to resort to other sources, such as the au-thors' personal webpages, to resolve practically all doubts.

Table 1. List of terms used in ISI and LISA

CLASSIFICAT* LANGUAG*
CLASSIFICAT* SYSTEM*
DESCRIPTOR LANGUAG*
DOCUMENT* CLASIFICATION
DOCUMENT* ORGANIZATION*
DOCUMENT* LANGUAG*
INDEX* LANGUAG*
KNOWLEDGE ORGANIZATION
KNOWLEDGE STRUCTUR*
LIBRAR* CLASSIFICAT*
SUBJECT HEAD*
THESAUR*

Table 2. Search codes in the ISOC database

Clasificación	Contenido genérico	Búsqueda completa o parcial
200100	Scientific output	200104 (Basic and applied research)
200200	Documentation and Information Policies	200200 al 200299
200300	Information Resources	200300 al 200399
200400	Information Analysis	200400 al 200499
200500	Information Management, Information torage and Information retrieval	200500 al 200599
200600	Información Indistry and Tecnology development	200604 (Automatic Indexing)
200700	Library System	200700 al 200799
200800	Archives and Museums Documentation	200802 (Archives Management)
200900	Information Management	200900 (Information Management);
		200901 (Services Planning)
201001	Docencia	Restricted to Knowledge Organization

Table 3. List of terms used in REBIUN and RUECA

```
LENGUAJE* DOCUMENT*
LENGUAJE* DE INDIZ*
LINGÜÍST* DOCUMENT*
CLASIFICAC* BIBLIOGRAF*
SISTEMA*DE CLASIFICAC*
CLASIFICACION* DE BIBLIOTECA*
CLASIFICACIÓN* DE LIBRO*
INDIZACION
CLASIFICACION
ESTRUCTURA* CLASIFICAT*
ESTRUCTURA* CONCEPTUAL*
ENCABEZAMIENT* DE MATERIA*
LISTA* DE ENCABEZAMIENTO*
ORGANIZACIÓN DEL CONOCIMIENTO
THESAUR*
TESAUR*
```

After normalizing the records using Procite, we generated the indexes needed for the analysis of research output. These were exported to Excel to generate the corresponding tables and the graphics.

– Who researches and publishes?
– How much research is published?
– Where is the research published?
– When was the research published?
– What was the topic of the research and publication?

3. Results

The heterogeneous nature of the collections comprising the sample of our study meant that different types of documents had to be analyzed in an isolated manner at first. That is, we studied the articles, the monographs and the dissertations separately. Each group has its own characteristics, which may affect not only the type of study done, but also the material dealt with. For instance, the descriptive model of the monographs does not indicate the origin or the institutional setting of the authors' work (this information does not usually appear in a book). Similarly, monographs have particular features we will discuss below. The final section under Results includes an appraisal of the data altogether, which would reflect the general patterns of Spanish research in the field of KO. The contribution of each database to the sample used in the present study in the number of references is shown in Table 4.

Table 4. Records in the databases consulted.

Data base	Records
ISOC	171
LISA	51
ISI	17
REBIUN/RUECA	145
TESEO	15
Total:	399

Knowl. Org. 31(2004)No.3
M. J. López-Huertas and E. Jiménez Contreras: Spanish Research in Knowledge Organization (1992-2001)

141

LISA gave a greater yield of pertinent references at first, but was largely reduced later on when we detected duplications in the ISOC and ISI. It was therefore decided to leave in LISA only those records that were not repeated in either of the other two databases.

3.1 *Authors and the quantification of their output*

When articles, monographs and theses were analyzed (separately, as explained earlier), we obtained the following results.

3.1.1 *Authors of articles and their production*

There were 201 article authors identified, producing 330 articles in national and international journals in the decade 1992-2001. Only a small number published more than four articles in the period of study. A summary of the most productive authors is given in Table 5.

The total number of authors – 201 – might seem high at first, but in view of the production of every single author, we see that about three-fourths have made only one contribution over the decade of study. This sort of testimonial presence might mean that these authors study other subjects related to the area of LIS, and that they sporadically touch upon subjects related with KO. However, it may also point to a lack of research activity. It would be interesting to cross these data with the institutional affiliations of the authors to detect any association with low output.

Table 5 clearly shows that only 18 researchers of the 201 identified can be considered "productive" in developing and diffusing KO in our country. This stands as 9% of the total. Table 6 shows the number of authors of articles in comparison with their production by percentages.

Table 5. Most productive authors of articles.

Esteban Navarro, M.A.	9
Garcia Marco, J.	9
Izquierdo Arroyo, J.M.	9
López Alonso, M.A.	9
López-Huertas, M.J.	9
García Gutiérrez, A.	6
Gil Urdiaciaín, B.	6
Martínez Méndez, F.J.	6
San Segundo Manuel, R.	6
Currás, E.	5
Llorens, J.	5

Moreiro González, J.A.	5
Moreno Fernández, L.M.	5
Tramullas Saz, J.	5
Caro Castro, C.	4
Martín Pradas, A.	4
Rodríguez Muñoz, J.V.	4
Velasco, M.	4
Autores con 3 publicaciones	6
Autores con 2 publicaciones	25
Autores con 1 publicación	152

Table 6. Number and percentage of authors and publications

No. of authors	% of authors	No. of articles	% of articles
5	2,5	45	13,6
4	2	24	7,3
5	2,5	25	7,6
4	2	16	4,8
6	3	18	5,4
25	12,4	50	15,2
152	75,6	152	46,1
Total: 201	100%	Total: 330	100%

It is thus apparent that 9% of the researchers identified are responsible for 33.3% of total output in articles. The most productive authors (5, with 9 each) show a Lotkian distribution, with a reduced percentage concentrating the bulk of output. Also apparent upon comparison of these authors with their centers of production is that the active Spanish researchers come almost exclusively from the university setting. It is interesting to note that professionals who are active in Documentation centers or services are not identified as authors in the area of KO, although they would be in other LIS-related subjects.

3.1.2. *Authors of monographs and their production*

The production of monographs is lower than that of articles. The 141 total authors produce a total of 278 books. The idiosyncrasies of this group led us to deal separately with the individual authors and the collective authorships. From the research standpoint, individual contributions are more significant, but we did not wish to ignore the work destined to the production of tools for the processing of information, or normative work on elaboration (the subject matter of most collective publications), as there are interrelations of feedback.

The group corresponding to the individual authors is represented by 100 researchers who produce a total of 116 monographic works. This means that 70.5% of the total authors are behind 41.7% of the works. Productivity can be broken down as shown in Table 7. This Table, like Table 5, shows that that few authors have more than two publications on KO from the period of study. Only 11% of authors produced two or more books. It is interesting to note that many of these books are manuals, or are specially geared toward teaching, though others cover more general aspects of KO.

Table 7. Individual authors of monographs and their output.

Benito, M.	4
Currás, E.	4
García Marco, J.	4
Moro Cabero, M.	3
Arana Montes, M.	2
Chacón Pérez, E.	2
Gil Urdiciain, B.	2
Ortego de Lorenzo Cáceres, P.	2
Pastor, A.	2
Rey Rocha, J.	2
San Segundo, R.	2
Zapico Alonso, F.	2
89 autores	1
Total authors: 100	**Total: 116**

The number of collective or institutional authors of monographs is 41 (29% of the total monograph authorship), with a contribution of 116 publications (41.7% of the total). The most productive groups are shown in Table 8.

Table 8. Most productive collective authors of monographic works.

	21
CINDOC	
Universidad Carlos III	19
Asociación Española de Normalización y Certificación Madrid	8
Universidad Complutense de Madrid	7
Centro de Estudios y Experimentación de Obras Públicas	6
Consejo Superior de Investigaciones Científicas	5
Universidad de Sevilla	4
Instituto de la Mujer España	4

España. Dirección General del Libro y Bibliotecas	3
Estudio de Técnicas Documentales Madrid	3
ISKO	3
Grupo de Trabajo de Archiveros Municipales de Madrid	2
Collegi Oficial de Bibliotecaris Documentalistes de Catalunya	2
Instituto de Consumo	2
27 autores colectivos	1
Total authors: 41	**Total: 116**

There is clearly a difference from the group of individual monograph authors. Nearly all these publications focus on the systems of classification, thesauri, subject headings, and norms for their elaboration, while few involve their didactic presentation. It is logical that CINDOC (Centro de Información y Documentación Científica) is at the lead, because it is highly active in the production of thesauri and the organization of some instructional courses. The publications stemming from the professional practice, then, seem quite different from those proceeding from the university sector when it comes to KO, something that is not necessarily seen in other LIS subject areas.

Finally, we made a joint list of the most productive individual monograph authors and article authors to get an integral view of the group in terms of total productivity, seen in Table 9.
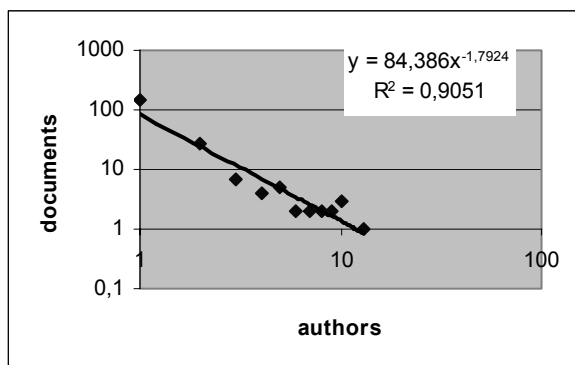
Table 9. Most productive authors of monographs and articles.

	Artí-culos	Mono-grafías	Total
García Marco, J.	9	4	13
Izquierdo Arroyo, J.M.	9	1	10
López-Huertas, M. J.	9	1	10
López Alonso, M. A.	9	1	10
Esteban Navarro, M. A.	9	0	9
Currás, E.	5	4	9
Gil Urdiciaín, B.	6	2	8
San Segundo , R.	6	2	8
García Gutiérrez	6	1	7
Martínez Méndez, F. J.	6	1	7
Martín Pradas, A.	6	0	6
Moreno Fernández, L. M	5	1	6
Moreiro González, J. A.	5	0	5
Llorens, J.	5	0	5
Tramullas, J.	5	0	5
Benito, M.	1	4	5
Moro Cabero, M.	2	3	5
Velasco, M	4	0	4
Zapico Alonso, F.	2	2	4

	Artí-culos	Mono-grafías	Total
Rey Rocha, J.	1	2	3
Chacón Pérez, E.	0	2	2
Arana Montes, M.	0	2	2
Ortego de Lorenzo Cáceres, P.	0	2	2
Pastor, A.	0	2	2

Altogether, 26 authors can be considered as the most productive ones; this includes all the authors that have published more than one monographic work or article. It is worth noting that seven authors of monographs have no output in the form of articles, and are therefore not included in the databases consulted. Likewise interesting, however, is that the top producers of monographs are top producers in articles. Taken as a basis the complete data for production by authors, a Lotkian table can be calculated, the index –1,792 (See Table 10)

Table 10. Most productive authors according to a Lotkian diagram

This productivity is, nevertheless, tentative, given the mediocre correlation coefficient that has been obtained. This is confirmed by the fact that the K-S test, used in these cases, yields a deviation much higher than the maximum allowed: 0,512 being $D_{max} = 0,19$. The shortage of works population, below 400, is probably the main reason of these results. Even so and taken the mentioned precautions, two facts associated with these data can be noted: there are no very productive authors, and there are a considerable group of medium sized producers who are the ones that explain this surprising high Lotka index.

3.1.3. Authors of PhD Theses

Only 15 dissertations dealing with KO were identified in the TESEO database. This is a low proportion of the 270 dissertations seen for LIS in general

(Delgado López-Cózar, 2002, p.88). Nine centers were responsible for these dissertations, with the University of Valencia at the head, and Granada, Complutense and Carlos III tied for second place.

3.1.4. Most cited authors

The current design of databases consulted allowed us to study only international publications included in the Web of Science, as only they dispose of the information needed to carry out such a study. The most frequently cited articles on an international level are identified with the following results. Of the 17 publications identified, only seven were cited. The number of citations according to country of origin is given in Table 10. The number of citations from Spain is relatively high, as to be expected, yet Spanish output also has substantial international visibility, with two-thirds of total citations coming from abroad.

Table 10. No. of citations by countries

Spain	6
USA	4
France	3
Reino Unido	2
Noruega	1
Chile	1
Italia	1
China	1
Grecia	1

3.2. Institutional affiliations and city of origin of the authors

3.2.1. Authors by city of origin

The spatial distribution of the authors reflects the geography of Spanish research on KO. The results (Figure 1) of our search show the cities to be the most important.
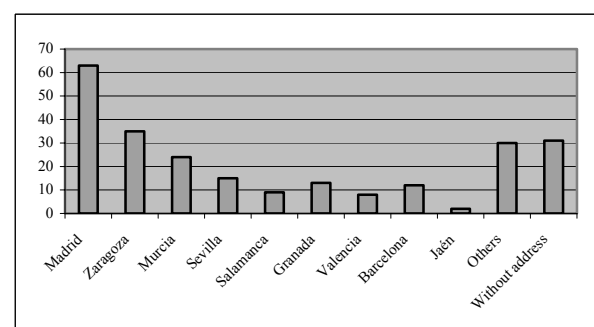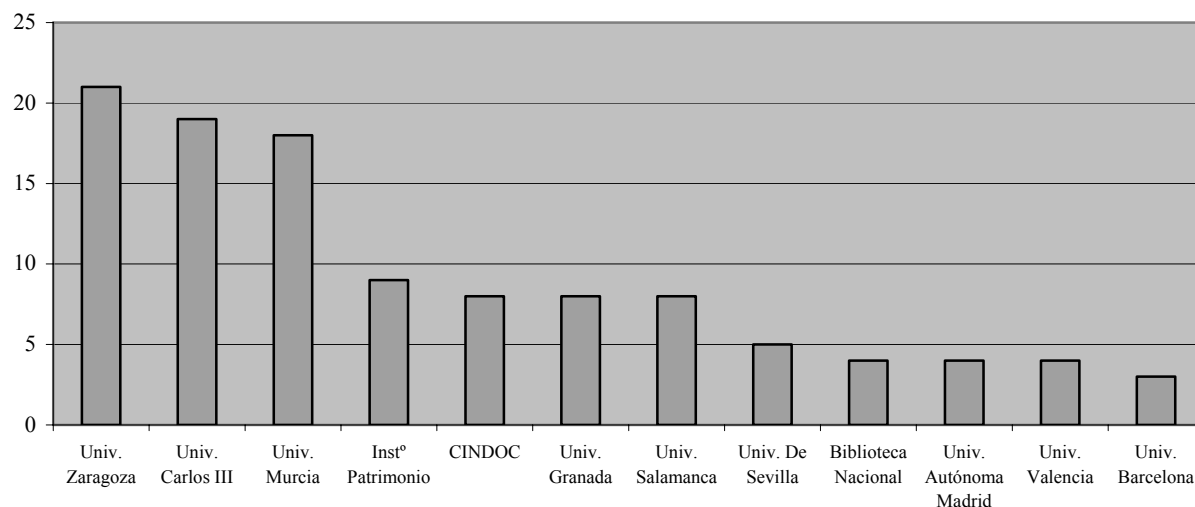
Figure 1. Authors by cities

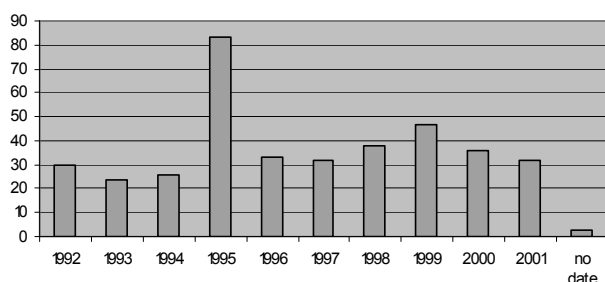Figure 2. Authors by institutional affiliation

3.2.2. Authors by institutional affiliation.

As seen in Figure 2, nearly 80% of researchers with publications about KO are affiliated with a University. The Complutense of Madrid and the University of Sevilla have a notable presence here, largely due to the activity of their respective libraries in relation with the publication of material Headings. The rest of the centers that are not LIS schools owe their presence mainly to the publication of Heading lists and Thesauri (such is the case of the CINDOC and the *Instituto del Patrimonio*) or to the publication of monographs that are actually courses on some documentary language.

3.3. Chronological evolution of productivity

In general, a growing trend is seen over the decade studied, with a remarkable peak in 1995, and a more discrete one in 1999 (Figure 3). The reason for the pronounced increase in 1995 would be the publication of the conference papers of the first two ISKO-Spain conferences.

Figure 3. Production by year.

Though held in Madrid in 1993 and 1995, actual publication of both took place in the same year, 1995. The appearance of the journal SCIRE, dedicated mainly to topics related with KO, also occurred at this time. The year 1999, similarly, saw the publication of the acts of the ISKO-Spain conference held in Granada that year. Other years show a fairly steady flow of output, with a slightly increasing trend. In view of these results, we must give due credit to the ISKO conferences for their impact.

The evolution of dissertations on KO was on the rise until 1996, then came to a standstill (Figure 4). It is interesting to note that the most productive authors in this field (Table 9) made their dissertations in areas other than Knowledge Organization.

Figure 4. Ph.D dissertations by year

3.4. Research topics

All the records included in this study were indexed, which facilitated a regrouping of the material to observe the subject matter. The results of this analysis will be discussed by sectors (articles, monographs and dissertations).

3.4.1. Thematic distribution of the articles

The variety of subject matter is evident, though it appears greatly summarized into categories to facilitate its representation. It is shown in Figure 5. The fact that most authors dealing with these topics are from the university setting, not from libraries or documentation centers, is noteworthy. Indeed, the focus of contents is largely theoretical, regardless of

the specific topic addressed, which might explain the absence of authors from the non-university setting.

When we reduce this sample to major groupings to improve the visualization of results (Figure 6) we see that the thesaurus stands out as the subject matter of 36% of publications. These thesauri can be broken down into more specific topic areas, as shown in Figure 7. The elaboration of thesauri can be identified as a main interest of Spanish researchers. It represents just over 20% of the total. Deserving mention is the manifest interest in specialized thesauri seen in the different areas of knowledge. If all these publications are put together, thesauri loom ahead as the topic of 42% of total output. More general subjects such as the use of thesauri in databases, or theoretical aspects of thesauri, show less presence.

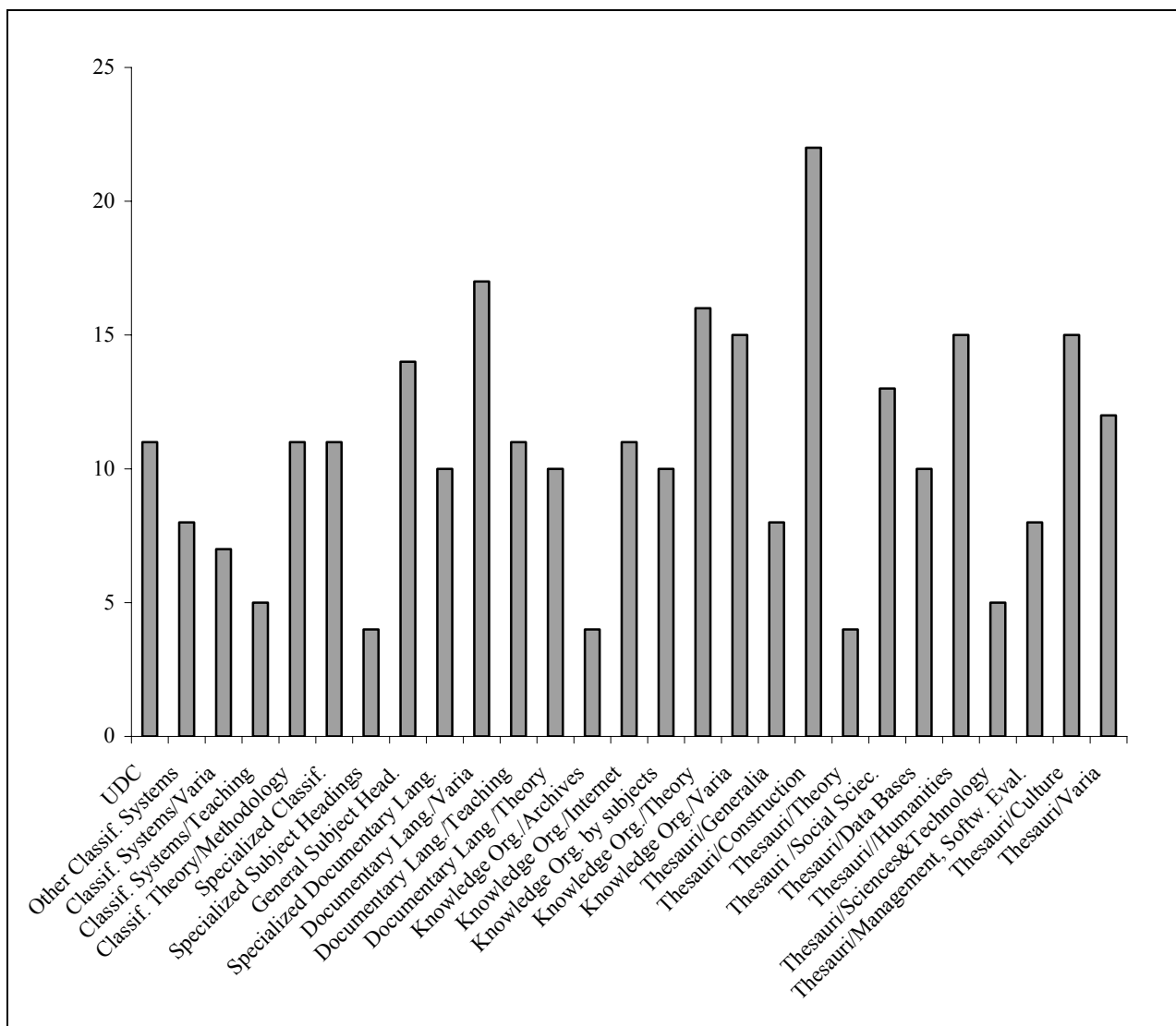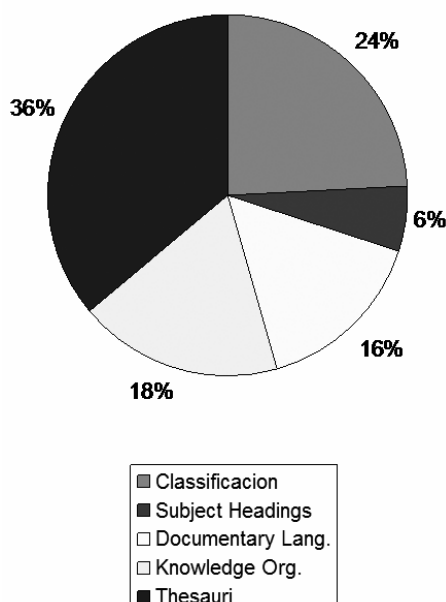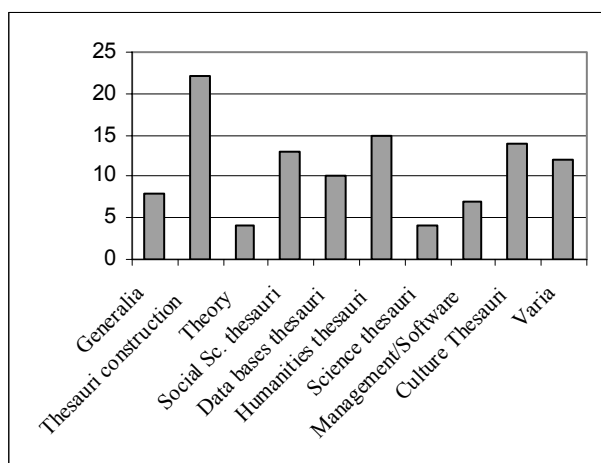*Figure 5. Thematic distribution of Knowledge Organization*

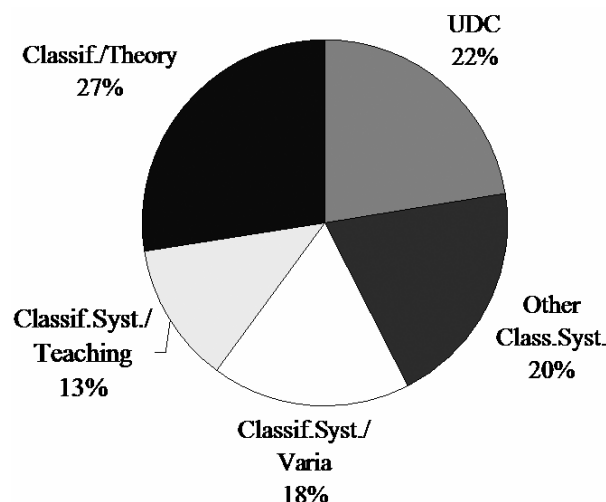*Figure 6. Major categories of Knowledge Organization (LISA, ISI, ISOC)*

It is interesting to note that both in theory and in the other subject areas included in the graph under "varia", systemic and cognitive approaches predominate.

Figure 7. Major topics in thesauri

The second most important thematic group is that of classification. While the focus of attention of a number of researchers it seems to have fallen behind since 1994 (Sánchez Casabón & García Marco, 1995). Figure 8 gives the picture in greater detail.

Figure 8. Thematic contents of Classification

Studies involving the UDC and classification theory can be seen to make up about half of the sample. Theoretical studies are substantially more important here (27%) than in the context of thesauri (4%), underlining "classification" as a very important nucleus of research activity in knowledge organization in Spain. Topics related with "teaching" are also quite important, which may indicate a higher degree of difficulty of the classifications with respect to other languages such as the thesaurus.

Figure 9 shows the importance of theoretical approaches in documentary languages and in KO (similar to what was seen earlier in the case of classification). Didactic aspects of documentary languages are important, as they are in other areas of LIS. The contents in the "varia" category refer mainly to general studies of the subject. KO on the Internet, and the rise of archives as one of its settings for application, are also noteworthy.

3.4.2. Thematic distribution of monographs

The subject areas of monographs present the peculiarity of belonging to two major groupings: the structures of knowledge themselves (classifications, subject headings and thesauri) which we call "texts" (44%); and studies of any aspect of KO and of any documentary language, which we call "studies" (56%). Figure 10 shows the breakdown. The group "texts" is included because we are dealing with publications, it has a considerable volume, and because it is an evidence of the professional contribution to

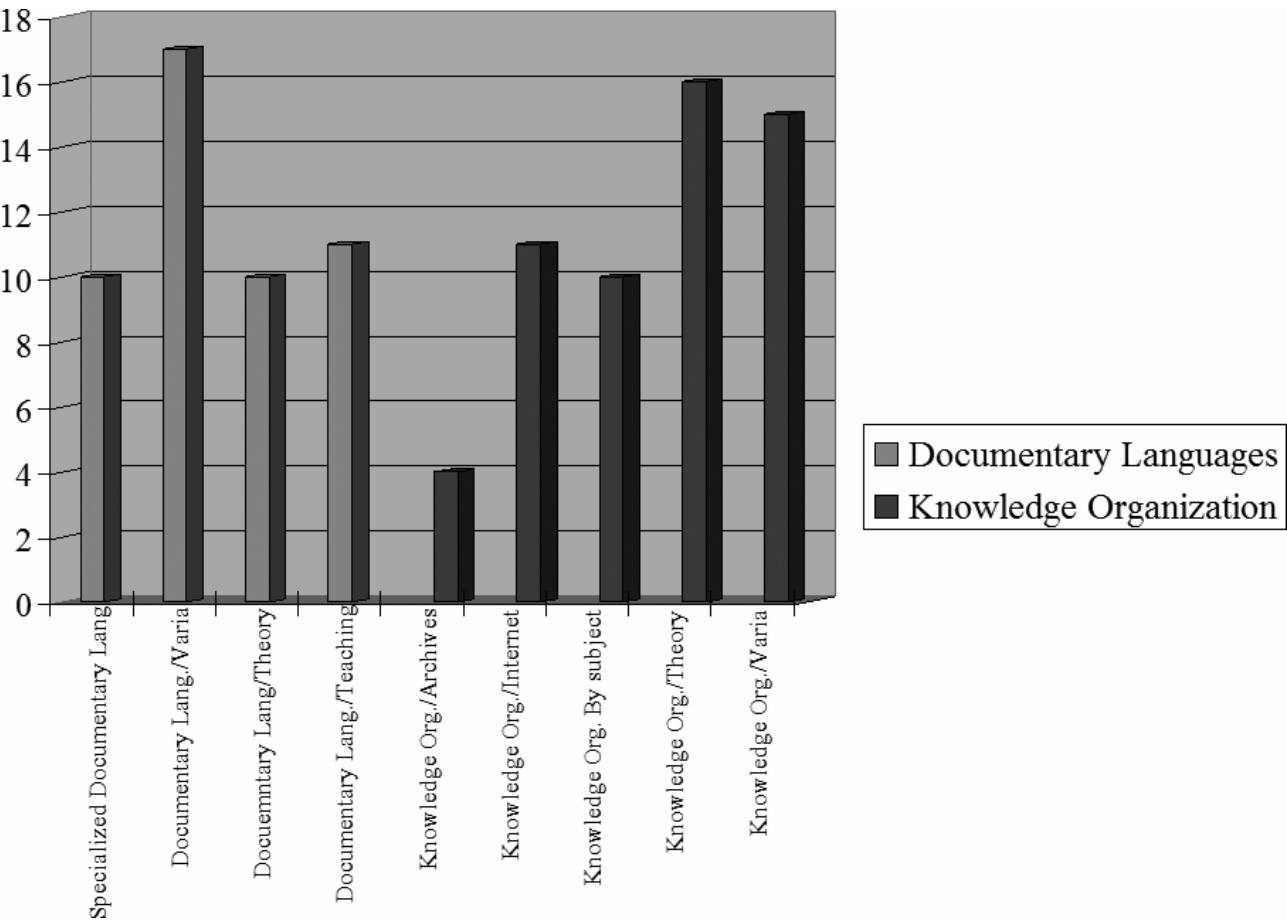*Figure 9. Contents of Documentary Languages and Knowledge Organization (ISI, LISA, ISOC)*
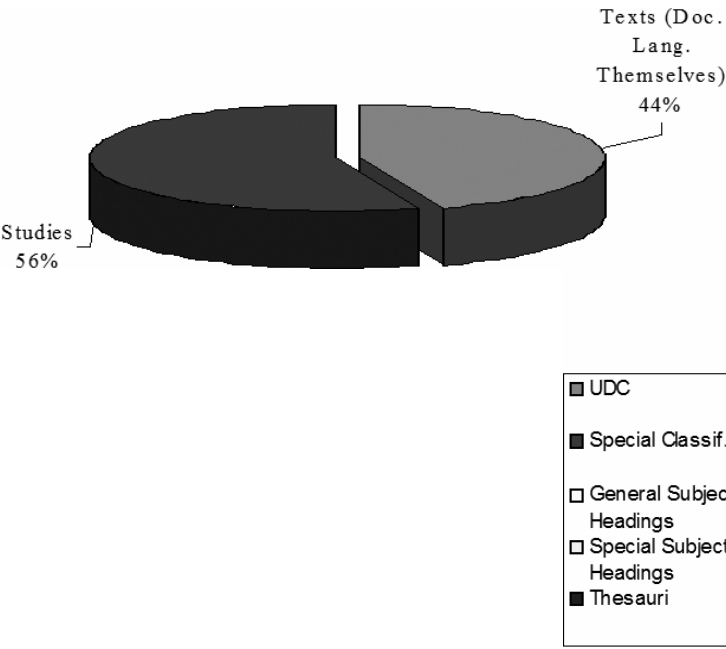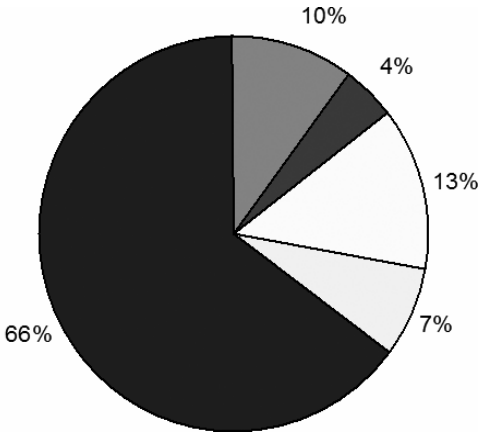

Figure 10. Monographs

Figure 11. Breakdown of monographic works (texts)

this sector. The most prevalent subject area, at the core of 66% of the monographic "texts" is the thesaurus, as seen in Figure 11. The second most important are the general subject headings, followed by UDC and special subject headings.

The group of monographic "studies" presents the topics seen in Figure 12. Again, the most prevalent are the thesaurus, followed by documentary languages, the UDC and other classification codes, and Knowledge Organization. Manuals on any of these

topics are also important in the general account of publications. The PhD Theses published are mostly dedicated to thesauri, classification, documentary languages, semantic relationships, control of vocabulary and documentary linguistics, in that order.

The topics of all the sources consulted are given in the graph of Figure 13. Again, the most frequent topic is the thesaurus, followed, at considerable distance, by classifications, documentary languages and KO. The *UDC* and the Subject Headings are last.

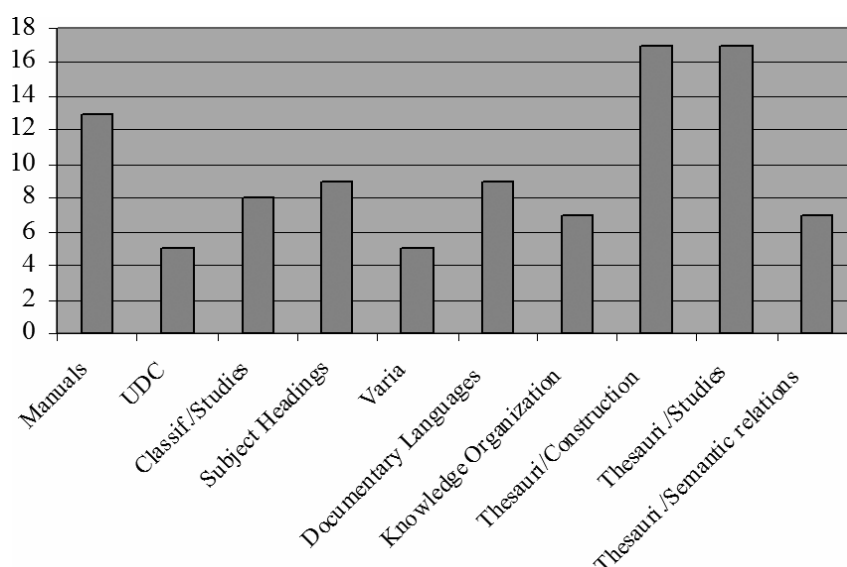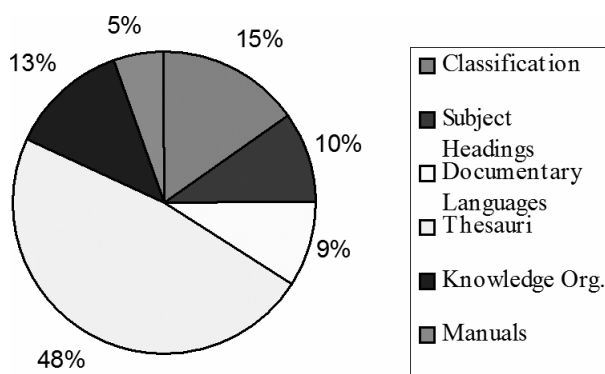*Figure 12. Topics of monographic works (studies)*

Figure 13. Most relevant topics in knowledge organization (articles and monographs)

Conclusions

In view of the results presented here, we may conclude that a small nucleus of authors is largely responsible for the development of this discipline in Spain, whereas the pool of other contributors has a

testimonial presence. This is seen very clearly in the production of articles, where the sporadic presence of authors stands for 75% of the total. It would be interesting to study the total output of these authors to see what other subjects are in the realm of their interest and the relationships they might share with KO. No collaborations are registered in the group of articles that make up our sample.

The international presence of Spanish research is very scarce, and its influence limited, given that only five publications out of the 17 identified are ever cited. The number of citations is symbolic.

The presence of collective or institutional authorships in the case of some 30% of the total monographs stands out as a distinctive phenomenon in the overall panorama. The bulk of these publications deals with conceptual structures themselves (thesauri, classification systems, subject headings, norms for their elaboration) and courses on any of these languages. The collective authors represent information and documentation centers and the libraries of vari-

ous organisms. For this reason we conclude that this is the best way to incorporate the professional LIS sector into the area of publications on KO.

The limited visibility of Spanish KO researchers showed in bibliometric studies devoted to LIS suggests that even authors who have outstanding productivity were not identified in these studies. The fact that the topics related with knowledge organization probably appear in these works subordinated to other categories such as the storage and retrieval of information, bibliographic control, etc., veils the references in databases specific to these subject areas. This finding is described by other authors in the international context (Hjørland, 2002).

Clearly the most important locus of activity is Madrid, followed by Zaragoza, Murcia and Granada. It is natural for Madrid to head the list as it holds a number of centers belonging to the different universities and schools of Library and Information Science, as well as other institutions. The University of Zaragoza is followed by Carlos III of Madrid as the institution producing most research, then by the Universities of Murcia and Salamanca. Intermediate positions are occupied by the Universidad Complutense and by two non-academic institutions: the CINDOC and the *Instituto del Patrimonio*. Their publications revolve around documentary languages, norms and courses.

As a general conclusion we can state that the Spanish universities are the centers of the universe of publication. The rest of the centers do not generate research in the strict sense, but contribute with professional manuals, documentary languages, etc. For instance, the CINDOC has limited output in KO as opposed to other LIS topics, and much of it consists of thesauri and courses.

Average output is quite low if we bear in mind that the top contributors have published nine articles within the decade of study. Still, a slightly positive trend is seen, suggesting that KO is on the road to consolidation as a research area. The ISKO conferences would appear to be largely responsible, directly or indirectly, for our output and interest in this area, as productive peaks can be traced to the meetings in 1995 and 1999.

There is quite a variety of specific topics in the material published, though the predominating categories are Thesauri, Classification, Knowledge Organization, Documentary Languages and Material Headings.

Theoretical aspects draw the most attention, seen above all in the contexts of Classification, Knowl-

edge Organization and Documentary Languages. Cognitive and systemic approaches are the most frequent, but not the only ones. The use of other disciplinary models – including linguistics and, more importantly, Terminology – also contribute to the development of the area.

A cross-disciplinary inclination is reflected in the applications outside the professional setting. Medicine is one discipline where this is evident. Likewise, there is a close conceptual proximity between Knowledge Organization and knowledge management in the business sector.

The study of concrete structures is also a mainstay of publications. The thesaurus is the basic focus of 41% of the articles published. Bibliographic classifications make up 42% of the output, and the UDC is 22% thereof. The only clearly downward trend can be seen with respect to Subject Headings, in which there is decreasing research interest. Teaching is an important source of interest, above all when it involves classifications or documentary languages in general. Manuals conform to 5% of publications. Yet the elaboration of thesauri is one of the most popular topics for Spain's KO-related researchers.

Other recurring themes are the use of languages in databases, mostly thesauri, the organization of knowledge on the Internet, and aspects related with the structure, semantic relationship, descriptors and categories, represented as *Varia* or *Other* in the graphics.

In short, it can be said that Knowledge Organization is taking shape as a research area in our country, and exhibits a positive, rising trend. Still, in general terms, above all considering the lack of visibility on the international forefront, we need to publish more, and branch out more. Research is too limited to the Universities of Zaragoza, Madrid, Murcia and Granada.

References

Andersen, Jack (2002). Communication technologies and the concept of knowledge organization. A medium-theory perspective. *Knowledge Organization, vol. 29 (1)*, p. 29-39

Cano, V. Bibliometric overview of Library and Information Science in Spain (1999). *Journal of the American Society for Information Science, 50(8)*, pp. 675-680

Delgado López-Cózar (2002), *E. La investigación en Biblioteconomía y Documentación*. Gijón, Trea, 2002

Delgado López-Cózar, E. (2000). Diagnóstico de la investigación en Biblioteconomía y Documentación en España (1976-1996) *Revista de Investigación Iberoamericana de Ciencias de la Información y Documentación, 1*, pp. 79-93

Frías, J. A. y Romero Gómez, P. (1997) ¿Quiénes son y qué citan los investigadores que publican en las revistas españolas de Biblioteconomía y Documentación? . *Revista Española de Documentación Científica, 1*, pp. 29-53

Hjorland, Birger (2002). The methodology of constructing classification schemes: A discusión of the state of the art. In *Challenges in knowledge representation and organization for the 21st century. Integration of knowledge across boundaries. Proceedings of the 7th International ISKO ...* Ed. by Maria J. López-Huertas. Ergon, 2002, p. 450-456

Jiménez Contreras, E. (2002). La aportación española a la producción científica internacional en biblioteconomía y documentación: balance de diez años (1992–2001). *Bibliotéconomia y Documentaciò, 9*, p. 979

Jiménez Contreras, E. y Moya Anegón, F. (1997). Análisis de la autoría en revistas españolas de Biblioteconomía y Documentación, 1975-1995. *Revista Española de Documentación Científica, 20/3)*, pp. 252-267

López-Huertas, M. J (1991). Lenguajes documentales: Terminología para un concepto. *Boletín de la ANABAD, (1991), XLI, no. 2*, pp. 171-188

López-Huertas, M. J. (1996). Thesaurus structure design: a conceptual approach for improved interaction. *Journal of Documentation, 52, 2*, pp. 139-172,

Moya Anegón, F. y Jiménez Contreras, E. (1998). Research fronts in Library and Information Science (1985-1994) *Scientometrics, 42(2)*, pp. 229-246

Moya Anegón, F. y Jiménez Contreras, E. (1999). Autores españoles más citados en Biblioteconomía y Documentación. *El Profesional de la Información, 8(5)*, pp. 28-29

Pérez Álvarez-Osorio, J. R. (1997). Cobertura temática y procedencia institucional de los artículos publicados en la Revista Española de Documentación Científica en sus veinte años de existencia. *Revista Española de Documentación Científica, 20 (4)* p. 290-298

Ruiz Pérez, R.; Delgado López-Cózar, E.; Jiménez-Contreras, E. (2002). "Spanish personal name variations in national in international biomedical databases: implications for information retrieval and bibliometric studies". *Journal Medical Library Association, vol. 90*, no. 4, p. 411–30.

Sánchez Casabón, A. I. y García Marco, Javier (1995). La investigación sobre el análisis de contenido y los lenguajes documentales en las publicaciones periódicas españolas de Información y Documentación (1982-1994). *Revista Española de Documentación Científica, 18 (2)* p. 155-171