154

Knowl. Org. 38(2011)No.2
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

Designing a Thesaurus to Give Visibility to the Historical Archives in the Archivo del Reino in Valencia

Vicent Giménez-Chornet* and Mercedes Escrig-Giménez**

* / ** Departamento de Comunicación Audiovisual, Documentación e Historia del Arte, Universidad Politécnica de Valencia (UPV), Escola Tècnica Superior d'Enginyeria Informática Universitat Politècnica de Valencia, Camí de Vera, s/n, 46022 València, València, Espanya / Spain, * <vigicho@har.upv.es> ** <maesgi@har.upv.es>

Vicent Giménez-Chornet is Professor of Documentation Studies at the Polytechnic University of Valencia in Spain. He is a member of the Documentation Committee of the Spanish Association of Standardization and Certification (AENOR) and President of the Official School of Librarians and Documentalists for the Valencian Region. He has worked in the Archivo del Reino of Valencia and in administrative archives. His field of research is information retrieval in archives and the management of electronic documentation.

Mercedes Escrig-Giménez is Associate Professor of Documentation and Librarianship at the Polytechnic University of Valencia in Spain and the Head of Archive Administration at the Valencian Regional Goverment Department of Health. She has an M.A. in The History of the Book from The School of Advanced Studies at the University of London and has been Head of the Archivo del Reino in Valencia, Head of the Information Department of the Regional Valencian Government President's Office, and Head Librarian of the Cervantes Institute in London. Her field of research is the study of accessibility, visibility, and procedures for archival documents.

Giménez-Chornet, Vicent and Escrig-Giménez, Mercedes. **Designing a Thesaurus to Give Visibility to the Historical Archives in the Archivo del Reino in Valencia.** *Knowledge Organization, 38(2)*, 153-166. 29 references.

ABSTRACT: This article describes the process of constructing a thesaurus based on original historical documents located in the Archivo del Reino de Valencia, one of the most important historical archives in Spain. After examining precedents, we discuss the methodology used and its application. The aim of the project was to provide visibility to information contained in these historical documents which would otherwise have remained inaccessible and to do so in the clearest, most rigorous, and most useful way possible for both the specialized user and the general public. The use of information technology as a management tool is not as common in archives, especially historical archives, as in other documentation areas such as libraries. A pilot project was therefore set up to create an intranet support for managing four specific historical archives containing documents dating from the 13th to the 20th centuries. The most important part of this project was that of exploring the viability of constructing a thesaurus to become part of the automated program for archive description and checking.

Knowl. Org. 38(2011)No.2
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

155

1.0 An historical overview

In medieval times, vocabulary to identify documents usually consisted of a list of words in alphabetical order written at the beginning of a register into which administrative documents were then copied. This enabled the user, usually part of the organization concerned, to find where the document could be found. The user's need was to be able to find the relevant document rapidly, and the efficacy of his search depended on the abilities of the administrators of the archive. The terms they used were of two kinds: names of institutions and organizations, and names of people. This can be seen in the surviving documentation, in our archive, for example, in the volumes of the series entitled Real Cancillería del Archivo del Reino de Valencia. These volumes, arranged in various series, were begun in about 1419. On the opening pages can be found the lists of index words, arranged alphabetically, which refer to the page where the document of the Valencian chancellery can be found. The close relationship between the users or interested parties and the administrative organization was of prime importance when indexing such documents. As the priority was to determine where a document relating to a particular person, institution or organization was located, subject terms were not usually necessary, and the few cases where they were used were generally only for documents specifying standards, as these were concerned with specific issues.

The beginning of a standardized vocabulary was in the field of public or private law, where the subject matter was also important. In the case of public legal documents, the information given would be a) the type of proceeding, together with b) basic classification of the issue in question (appeal, request, gift, will, etc.). In the case of private cases, together with the identity of the interested party or parties, there would be an indication of the category, normally standardized in these types of documents (purchase or sale, will, inheritance lines, marriage announcements, etc). However, this standardized subject vocabulary was never used to create alphabetical lists that referred to a document, but to relate the person to the matter concerned. Such lists were therefore created based on personal or institutional names, or else were chronological.

In the 19th century, index cards as finding aids began to be used. By indexing documents on separate cards, descriptions began to be made using normal language. To arrange the cards in order, however, standardized terms were used in which subject terms were combined with names of persons or institutions, with the term used written on the upper part of the card so the cards could be arranged alphabetically.

2.0 The need for a thesaurus and appropriate information retrieval strategies

Unfortunately, in our case, few archives arrive as the result of some kind of standard procedure. Archive documents are all too frequently treated as a collection of material objects. Traditionally, the tools used by archive professionals to describe such collections have been inventories and catalogues. The introduction of information and communication technologies (ICT) in archives, however, is changing, if not the structure of these tools, at least the way in which they are presented, and the way users access them. The use of software designed to organize documents, accompanied by precise standards to process the information, therefore, provides new approaches providing visibility to the contents of archive documents. Among various possibilities, we examined the use of specific language components, that is, the development of a thesaurus designed specifically for analyzing original historical documents.

The problem of information retrieval in historical archives has been dealt with in the professional literature in recent articles from several different points of view. The proposals made by M.L. Nielsen (1999) at the IV Congreso ISKO-España emphasized the point of view of the systems user in the recovery of information and emphasized the usefulness of thesauri when full text recovery is not possible. This is the case with the collections in historical archives where issues of paleography and languages make them inaccessible for many users. These are the issues that we wished to deal with in our project. A useful definition of the basic relationships that typify a traditional thesaurus, its relational semantics, focusing on the hierarchical relationships, and presenting thesauri as "a tool capable of representing the complexity of knowledge and of the semantics of terms" can be found in Mazzochi, F. et al (2007, 198).

The problems of languages which are difficult to access, as in our case with Latin and Catalan, are discussed by Schenkolewski-Kroll and Tractinski (2006), who present the construction of a thesaurus in Hebrew as a solution when it stresses what is significant, for example, names of persons and corporate bodies. They also propose the following solution to synonym problems: "when persons who Hebraicized their name appear in various documents by different names, 'use'

156
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

Knowl. Org. 38(2011)No.2

and 'UF' references are used to list all material under one of the names. This thesaurus is arranged by the Hebrew alphabet" (Schenkolewski-Kroll & Tractinski 2006, 89).

A group of researchers from the universities of Athens and the Peloponnese proposed adopting a methodology that involves the construction and use of ontologies (Katifori et al 2008). The objectives described by these authors were particularly relevant, as they matched our own, which led to the conclusion to construct a thesaurus in-house not only for researchers, but also for the general public. They noted, for example that "This work is based on user study of historical information retrieval methods in order to create a set of tools for the context of historical archives, which will facilitate historical data storage, management and visualization" (Katifori et al 2008, 32). After explaining the methodological issues involved in the ontology they developed, they conclude: "This work proposes a toolset for modeling and visualizing historical information, aiming to assist researchers in studying historical material; the toolset makes special provisions for tracking and visualizing entity evolution which is an important task for historical research" (Katifori et al 2008, 37).

Schenkolewski-Kroll and Tractinski (2006) is a useful examination of how these issues have been dealt with from the perspective of the Israeli archives. They examine the various options implemented according to the development of standards and technology from the 1980s to the present day and conclude (p. 79):

"archivists throughout the world are in agreement, or near agreement, on the manner in which the materials deposited in their institutions should be described and on the importance of achieving this with controlled systems using standards for their description. Moreover, they have successfully been able to make them part of automated systems. However, much effort is still called for in order to corroborate these theoretical assumptions, especially in relation to the use of controlled vocabularies and their incorporation into computerized systems for archival description and IR."

The experience of the Israeli archives is especially relevant to our case due to the specialized nature of the collections. They confronted structural problems similar to our own, such as (Schenkolewski-Kroll and Tractinski 2006, 88-89): "a) The depth of indexing, b) Changes in terminology and c) Names of persons."

The creation of an in-house thesaurus is recommended, despite the problems of methodology and standardization involved. Schenkolewski-Kroll and Tractinski reach the conclusion that this matter is far from being resolved, however, and that their research is intended as a starting point for future research. It is hoped that our contribution will further clarify possible solutions to enable visibility and access in information retrieval.

3.0 European precedents

3.1 Spanish precedents

The perception that information retrieval tools in HA are needed is not new. Madrid Local Administration Archivists, conscious that one of the main functions of archivists is to facilitate and make known the contents of the treasures in their historical archives, explained why archivists and archives need such a tool (Merlos-Romero and Villarreal-Mascaraque 2001 #): "One of the fundamental tasks of archivists is to facilitate knowledge of the archival collection, both as a whole and of the individual documents housed there … a thesaurus, as an archive language tool, enables such information to be found, and is the means which produces the best results both in describing contents and finding information." A purpose-built methodology applied to an historical archive and available to consult online was created by Patrimonio Histórico Andaluz and, although it is not a thesaurus, it is important for its subject relationships (Instituto Andaluz de Patrimonio Histórico 2008).

With regard to thesauri, when normalizing terms and concepts for information retrieval in historical archives, it is especially important to understand, for example, names of professions which have now disappeared and their equivalents in the present day, especially if we are working with original documents from the Old Regime. An example of using specific vocabulary as part of thesaurus to resolve this problem and other problems based on original documents and referring to a specific place, in this case the city of Zaragoza in Aragon, and at a particular time, in this case the 16th century (Santos-Lorite 2002), lead us to consider the suitability of this tool, although applied to a wider context.

The differences between taxonomy, ontology, and thesaurus have already been sufficiently documented in academic literature. A good list of works can be found in Arano (2005), though in our opinion rather theoretically, and these terms are also defined by Gil-

Knowl. Org. 38(2011)No.2
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

157

christ (2003). On the subject of standardization, a recent article published by researchers at the University of Alcalá de Henares discussed some of these issues, and proposed the possibility of establishing ontologies based on document descriptions. The group also emphasized the need to "standardize the description of the contents of documents in order to facilitate access to the information" (Sánchez-Alonso et al 2008 12). There is, however, little practical experience in this area, and where there is, the purpose-built thesauri set up by archivists are of a general nature, such as that of the group of Madrid Local Administration Archivists referred to above, and have been used for the most part on administrative archives and modern documents.

3.2 British precedents

In Britain, as early as 1998, a report for the National Council of Archives in the UK stated the need to (Sargent 1998, 3): "bring together appropriate representatives of the national repositories … and the Historical Manuscripts Commission should examine the desirability of standardizing name authority controls at national level." One of the projects of the National Archives, conjointly with the Calculation Centre of the University of London, and thanks to funds provided by the Heritage Lottery Fund, was to implement the so-called UK Archival Thesaurus, or UKAT (2003), in the National Archives Network. Large archives and archival systems participated in this project at state, regional, and university level. The system is now functional and is being used and maintained. Details of the methodology used, along with participants and planning, are described. A specific example of its practical application can be found in the portal of one of the participating networks, the AIM 25, or "Archives in London and the M25 area," where it is possible to navigate through the thesaurus as a way of consulting and accessing the inventories of the various participating archives (AIM25 2008). The thesaurus is presented hierarchically and alphabetically, and, moreover, in the context of a public website.

At the British Library, specific regulations, although not available to the general public, can be found at the BL Integrated Archive and Manuscripts System. "IAMS also allows for the description of people, families, corporate bodies, places and subjects that are somehow related to the archives and manuscripts described and enables the relationships between them and between the archives and manuscripts to be defined" (British Library 2009, 6). It was

also of interest to consult the standards and methodologies specifically applicable to place names, personal names, organizations, and in establishing subject headings.

3.3 French precedents

In France, a tool particularly relevant for historical archives was developed to standardize specialist terms. A working group of the Direction des Archives de France undertook the specific task of processing information in historical archives from the point of view of old place names. As will be seen below, the problem of equivalences or relationships between current place names and old ones is a problem that had to be solved when analyzing the information contained in original documents and was one of the principal arguments in favor of creating a specific tool for resolving this kind of problem.

The third edition of the "Thesaurus W, vocabulaire normalisé pour la description et l'indexation des archives administratives locales contemporaines" was published in 1997, and, between 1989 and 2000, various updates appeared. The preface indicates that "Thesaurus W has the function of standardizing the archives of all geographical areas, départamentales, communales and régionales" (Direction des Archives de France 2008a, 1). The Direction des Archives de France has recently published the fourth edition of this Thesaurus, now called Thésaurus pour la description et l'indexation des archives locales anciennes, modernes et contemporaines, (Direction des Archives de France 2009). This makes it an essential tool for describing the collections in these and other French archives. An indication of this can be found in the index of the modern collections in the municipal archives of Lyon, where it states: "the terms found in Thesaurus W of the Direction des Archives de France, ed. March 1997, have been used for articles later than 1960" (Cadieu 1997).

There are also cases, however, such as that of the archives of the Conseil General des Alpes-Maritimes, where a list of key terms was established without much standardization, depending on the archive concerned, although it can be consulted using screens which provide the user with at least some degree of navigation for research purposes. The structure of the list, however, is only accessible alphabetically. In contrast, The Archives de Cannes is one of the best examples in the French language of the use of a thesaurus in the descriptions of the contents of their archive (Archives de Cannes 2009)

French curators and their technical teams were among the first professional groups to create tools for terminological standardization, and it is perhaps for this reason that they have not always been made available as research tools on Internet websites. There is therefore both diversity and inequality in their practical application. The Direction des Archives de France, however, offers a great deal of documentation on its website on the problem of archive indexing and methodology. An information note on "Indexing: the current situation" begins by pointing out that "The management of French Archives is aware of the numerous requests related to archive description indexing, either on paper of electronic ... The purpose of this note is to describe the principles to be followed in order to standardize key terms" (Direction des Archives de France 2008b, 1-2)

Among the documents and lists of standards available to professionals, one document in particular (Péllisier and Motte 2003, 4), is related to the standardization of geographical names. It states that "the needs of researchers, professional historians or amateurs, and the professionals who have the task of indexing old documents which include old place names" were the main reason for setting up a working group. This thorough report includes in its numerous annexes detailed examples of each case of equivalences, geographical agglomerations and divisions, all following a system of codes which, as it was to include all of the French state, was of considerable complexity.

3.4 Summary

The English and French examples provide us with two different approaches to terminological standardization. In the French example, perhaps due to its long documental tradition, there are many and diverse practices in standardizing key terms. On the one hand, the central administration seeks to standardize and prescribe the setting up and use of this type of tool, making a range of documents and standards available, but on the other, we find very uneven practical applications. This can be seen in the fact that, in the archives of the various local administrations, non-standard initiatives are being developed, in Lyon for example, which respond to the indexing needs of archives and specific collections, as indicated above. We await, nevertheless, the "historical geo-nomenclature" to be finally published. The British example is the opposite case. The project was initiated based on the experience of using a general the-

saurus, that of UNESCO (1995), which was found to be inadequate when applied to the descriptions of the collections of the network of archives. The working group then developed and implemented a project which began by collecting proposals for new key terms from the participating archives by means of a website set up for that purpose. All of the key terms proposed were considered, and, if they were finally accepted, were incorporated into the framework of the structure of the thesaurus. This same structure is accessible to all, including the user. It is, therefore, clear that it is possible not only to index the documents and collections in an Archive with a controlled language, but also to optimize this task by using it to facilitate reader searches by means of a website, as is the case in the British example cited above.

4.0 Methodological issues

4.1 Initial considerations

Thesauri do not, of course, in themselves, resolve the problems of language or paleography. What the prototype examined here was intended to facilitate, rather, is access to the catalogue entry for the item, so that the researcher could be aware of the existence of a catalogue entry for a document which contains a specific type of information. Once the document has thus been made accessible, or "visible," it is for the user to apply their paleographical or language skills as appropriate.

Due to the unique nature of the documents in our archive, it was clear that no existing thesaurus could provide either the semantic fields or descriptors and their interrelationship that were required for our stated objectives. We, therefore, proceeded to create a purpose-built thesaurus (Escrig-Giménez et al 2006), designed both for analyzing the contents of documents and to enable the information available for a specific collection of documents to be found. The process had to take the following factors into account: 1) we were concerned with original archive documents from the 13th to the 20th century, 2) documents prior to the 18th century were written in Latin or Catalan only, 3) paleography skills would be required, and 4) the nature of the documents made knowledge of the institutions of the Old Regime at the time of the fueros (legislation applying to the territories of the old Crown of Aragón in medieval times, that is, the ancient kingdoms of Majorca, Catalonia, Valencia, and Aragon, before its abolition at the beginning of the 18th century) essential, along with the appropriate vo-

Knowl. Org. 38(2011)No.2
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

159

cabulary and knowledge of old place names. For these reasons, the working party was interdisciplinary, and included specialists in history, archives, documentation, IT, and paleography.

4.2 Initial procedures

We started from the principle that the methodology to be followed when setting up a thesaurus for an archive should allow us to create a tool that could be used both to index and to analyze specific documents in the archive. The aim was not to create a theoretical thesaurus that covered all aspects of human activity, but rather one that was oriented to producing an index for an archive or specific archival system. The creation of the three thesauri, one each for 1) institutions, 2) geographical place names, and 3) subjects, was carried out in the following stages

Three specialists were appointed, each in charge of a thesaurus, based on their suitability and subject knowledge. Each specialist initiated the process of selecting vocabulary using the most appropriate sources, from the archival documents themselves to printed or online documentation. Once the vocabulary was selected, the specialist began the selection of terms considered to be the most suitable for indexing the archival collection, and to construct the hierarchies and semantic relationships required. Beginning with written notes, the information was then entered into the automated program. The hierarchies and relationships were then revised for each thesaurus, paying special care not to leave out any "empty" vocabulary, or to reject any ambiguous vocabulary, either due to multiple meanings or coincidence with another vocabulary item already used.

4.3 Software, standards and maintenance

The software needed for the creation, management, and maintenance of the thesaurus had to be provided with the basic facilities, such as the ability to insert a term, create hierarchical and associative relationships, and insert application notes. This software was then to be integrated into the database: This was made possible by using Informática El Corte Inglés (2008). This was an essential requirement, as it enabled any changes made to the thesaurus, such as alteration to terms or changes to semantic relationships, to be entered directly into the database without the task of changing the registers of analytical descriptions.

The Archidoc software used was first used in the automation of The Archivo General de Indias in

Seville at the end of the 1980s. The strategies and results of this project were presented and published by Prof. P. Gonzalez (1998). The software was created in order to describe and link digitized images at a time when the ISAD (G) and ISAAR-CPF standards had not yet appeared. This software was acquired by the Archivo del Reino de Valencia in order to establish the parameters for the database for the creation of fields according to international standards for archive description, and to create links between them and the digitized images. The original contribution of the Archivo del Reino de Valencia was to add to this the facility for creating a thesaurus linked to this database.

Thirteen archival specialists were appointed to analyze the documents, and the resulting descriptions were then indexed. This in turn enabled the documents to be searched efficiently, as controlled vocabulary was used in the process of locating the most relevant documents. The subject thesaurus included 2,107 terms, the geographical thesaurus 1,450 terms, and the institutions thesaurus 1,260.

A thesaurus is, however, in a continual process of change, and there is, therefore, the subsequent stage of maintenance. This involved the following procedures: a) the specialist responsible for a specific thesaurus was also responsible for incorporating new terms, or altering relationships; b) if the archivists who described and indexed the documents was lacking a term, they had to submit the term to the specialist responsible for that thesaurus for acceptance; c) the specialist responsible analyzed the proposed new term in the context of the documents that produced the need for it, and then proceeded to establish if there is an existing term with this meaning, or if it was necessary to incorporate a new term; d) the archive descriptor was then informed of the decision adopted, i.e. a substitute term or a new term.

4.4 Gathering terms

Gathering terms for the thesaurus can be carried out in various ways, although, logically, priority in gathering terms was directly related to the size of the collection. In our case, the first and main one was by means of the descriptions of the archives that had already been made, based on the archival documents themselves. Published bibliographical and documental resources were also used. Subsequently, terms that appeared on existing index cards were used. If only series of documents were identified, provided that they were described, a basic vocabulary was obtained from these descriptions. If, however, as oc-

160

Knowl. Org. 38(2011)No.2
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

curred in many cases, only the series was identified without descriptions, sample documents had to be used to determine the vocabulary, and the series then described as soon as possible.

4.5 Selection of terms and creation of a hierarchy

Once the terms had been gathered, the next step was to analyze terms which were synonymous and decide which terms would be descriptors and which would not be descriptors as such, but would be incorporated into the thesaurus as terms from which the user would be referred to the term selected. Automated systems were essential in this task, as they facilitated checking the number of terms selected, and in checking for empty terms, that is, those terms for which at this stage of the creation of the thesaurus no use had been made. Depending on the number of terms, this task was carried out at the same time as constructing the hierarchy and semantic relationships in the automated system. If the number of terms was very great, manual classification by categories or subject areas was carried out in order to be able to select the most general terms rather than the more specific.

4.6 Bibliographical and documental resources

Bibliographical resources and reference works, such as Currás (1998), were useful throughout the entire process of creating the thesaurus. Such works provided us with references to other published thesauri (Rubio-Liniers 1999), which had some similarities to the collections in our archive. An increasing number of thesauri are being published (Archivo Municipal de Arganda del Rey 2002) on various subjects, which provided general orientation or resolved problems related to semantic issues. In addition, dictionaries, both general language dictionaries and those specifically related to our subject, were indispensable tools for solving problems which arose concerning definitions and synonyms.

Finally, certain current or historical documents, especially prescriptive documents such as laws or decrees, were useful. In order to find suitable terms, it was essential to consult those original legal documents that regulated the creation and functioning of an institution. One example was the Gaceta de Madrid, Ley sancionada por S. M. sobre jurisdicción en materias de Hacienda (His Majesty's law on jurisdiction on matters concerning income tax), published 23rd February 1850. Similarly, recourse was made to technical reports or documents produced by an ar-

chive itself, for example the Reglamento que por soberana resolución de este día se ha dignado mandar S.M. se observe en las oficinas de administración, Contaduría, Archivo, Juzgado y dependencias del Real Patrimonio de Valencia. Madrid: por Manuel López impresor de dicho Tribunal, 1829 (Regulations which His Majesty this day by royal decree has ordered to be observed in the administration offices of Accounts, Archives, Courts, and the Royal Heritage offices of Valencia. Printed by Manuel Lopez, printer to the court, in 1829). Some data recently published on official Internet websites, such as city council information (CIVIS 2008) was also of great assistance, provided care was taken as to the accuracy of the data and whether the website consulted was genuinely official by consulting the official sources of the public administration body concerned.

5.0 Types of terminology problems

5.1 Semantic relationships in official institutions

The main terminology problems which occurred when setting up this thesaurus were inherent in the nature of historical documents themselves. As they are original sources, and the result of the activity of administrating an institution, or activity generated by persons, it is common for the organizations producing them, or even geographical places, to be subject to changes or alterations. The relationship between organizations producing documentation can often be confusing, given that different eras (e.g. Old Regime versus the constitutional era) had completely different official legal structures. This becomes more complex when different types of regions that developed differently, often characterized by the use of different official languages, nevertheless, came to form a single national-state. The former Kingdom of Aragon, for example, used both Catalan and Latin as official languages. From 1707 onwards, however, as a result of the victory of the Kingdom of Castile over that of Aragon in the Wars of Succession of 1707 – 1714, Castilian laws were imposed on those of the Kingdom of Aragon. As a result, the official language from that moment on was Castilian, i.e. Spanish, for all administrative organizations, even though the population spoke Catalan. The end result is a group of institutions or organizations the names of which do not always have a precise translation in the region where another language is used, as an organization with the same characteristics does not exist there. In other cases, the same institution name has undergone

changes due to new legal structures, and, as a result, its functional characteristics have also undergone considerable alterations. In our case, for example, in the 18th century, the Intendente General had political, taxation, legal and military responsibilities. At the beginning of the 19th century, however, the political responsibilities were passed to the Governador Civil, the taxation responsibilities to the Ministry of Taxation (Hacienda), and the legal ones to the Ministry of Justice, only the military responsibilities remaining with the Intendente, a post now exclusively related to the Army.

5.2 Semantic relationships with geographical descriptors

Geographical names, whether place names, or the names of rivers, lakes, or mountains, etc. have also undergone changes over the centuries. Place names change either because they have been incorporated into other regions where a different language is used, and are therefore re-named in their new language, or else, mainly for political reasons, it is decided at some point to change the name. An example of the latter in our Archive is the case of the town of Xàtiva, in the Valencian region. It was named Xàtiva in the era of the Kingdom of Valencia. After the War of Succession, however, the Spanish king Felipe V decided to change the name to San Felipe, a name it retained for more than a century. It was then changed to Játiva when the Spanish rather than Valencian spelling was used as, from 1707, the official language was Castilian (Spanish). Finally, it has recently returned to its original name, as the statutes of the current Spanish constitution officially recognize the various different languages spoken in Spain.

In other cases, changes to place names are the result of historical developments. There are two common reasons for this. Firstly, one place may become depopulated, and its place name might become absorbed by another already existing place, usually more dominant for its social, economic or political status. In our archive, this was the case, with Alcosser, which disappeared in the 18th century as the result of a flood, and whose place name became absorbed into that of Alberic. The second type occurs when it is decided that a well-populated center should become a separate entity, thus producing a new urban place name. Our example is that of Algemesí which, in the 16th century, was separated from Alzira, and they are now two separate and important cities in the Valencian region. When setting up a thesaurus it was important, therefore, to check for names of places which are now depopulated but which were formerly urban entities generating documents and which, therefore, were to be included in the thesaurus.

The names of rivers and lakes, etc. produce similar problems, as their names may change or alter as a result of being incorporated into new regions, or for political reasons. There are also problems resulting from the fact that the regions bordering on them may call them by different names. Alternatively, the same river may be called by different names in the various regions it passes through. This is the case in our region, for example, of the Guadalaviar, which is called the Guadalaviar where it rises in the province of Teruel, but is called the Turia in the Valencian region, where it reaches the sea.

5.3 Semantic relationships with subject matters

Terminology problems related to the subject matter in archival documents arise from the nature of original documents themselves. It is clearly not the same thing to produce documents on taxation, which will then form an archive, as to produce a theory of taxation, or taxation manuals. For a thesaurus to be an efficient archival tool, it should identify the subject matter of the original documents. Some possible terminological conflicts should be noted. One of these is the difference between identifying an institution or body and the various components of which it is formed. For example, the "Publishers Association" and "The Monastery of the Company of Jesus" are organizations in Valencia, but "publishers" and "Jesuits" are subjects. This is an important distinction, as proper names are controlled by the "bodies" section of the thesaurus, whereas "publishers" or "Jesuits" by the subjects.

In other cases, a terminology problem may be due to the fact that a term with one meaning is used for different purposes. For example, "matrimony" from the religious point of view is a sacrament, but, in legal terms, is a civil act related to family law. In other cases, the same term, through the centuries, can change meaning. For example, the term "adultery" was formerly an offence, and related to "criminality," and, indeed, it is still used with this meaning in some cultures. In modern democracies, however, it has now lost this semantic aspect. The significance of this will be seen below.

162

Knowl. Org. 38(2011)No.2

V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

6.0 Solving the terminology problems

6.1 Semantic relationships with official institutions

Provided that the identity of the institution corresponds to a particular period of time within a single legal jurisdiction, there are no particular problems with semantic relationships. If, however, we need to establish a semantic relationship between institutions that arose or carried out their activity under different legal jurisdictions, then the semantic relationship becomes difficult to establish due to this change in legal jurisdiction. For example, before the 19th century, the Mayor is an organization that had legal and political responsibilities as president of the City Council. From the 19th century onwards, however, the post continued, but without the legal responsibilities, which are now with the local judiciary.

Our solution was to set up groups of institutions according to periods of legal responsibility, and which are then placed under a term that is not itself a descriptor, as it is not an institution. This term was also to have a visual component that differentiates it from terms which are, indeed, descriptors. For example, [ADMINISTRACIÓN LOCAL DEL REINO DE VALENCIA], where the brackets have been added to

indicate that this is not a descriptor, as illustrated in Figure 1. In this particular case, the purpose was to group together separate hierarchies of descriptor terms related to local administration for a historical period in the Kingdom of Valencia between the years 1239 and 1707 (Escrig-Giménez et al 2006).

In cases where the institutions correspond to different autonomous regions using a different language, one language was chosen for identifying the institution when it came to creating the hierarchy. This is possible if, in the other language, there is a similar name that does not present difficulties in establishing a semantic relationship. For example, among the local Valencian institutions there is a position called "MOSTASSAF" and in the Valencian thesaurus, therefore, it appears in this form. In the Spanish version of the thesaurus, however, for the same term in the same hierarchy, "ALMOTACEN" has been used, a local Castilian position with the same functions as that of the Valencian Mostassaf.

If, however, the institution has different characteristics that are not represented in the other language, its name in its original language was used, and a scope note inserted describing the characteristics of that institution so that it was understandable in the language being used for the thesaurus. For example, the
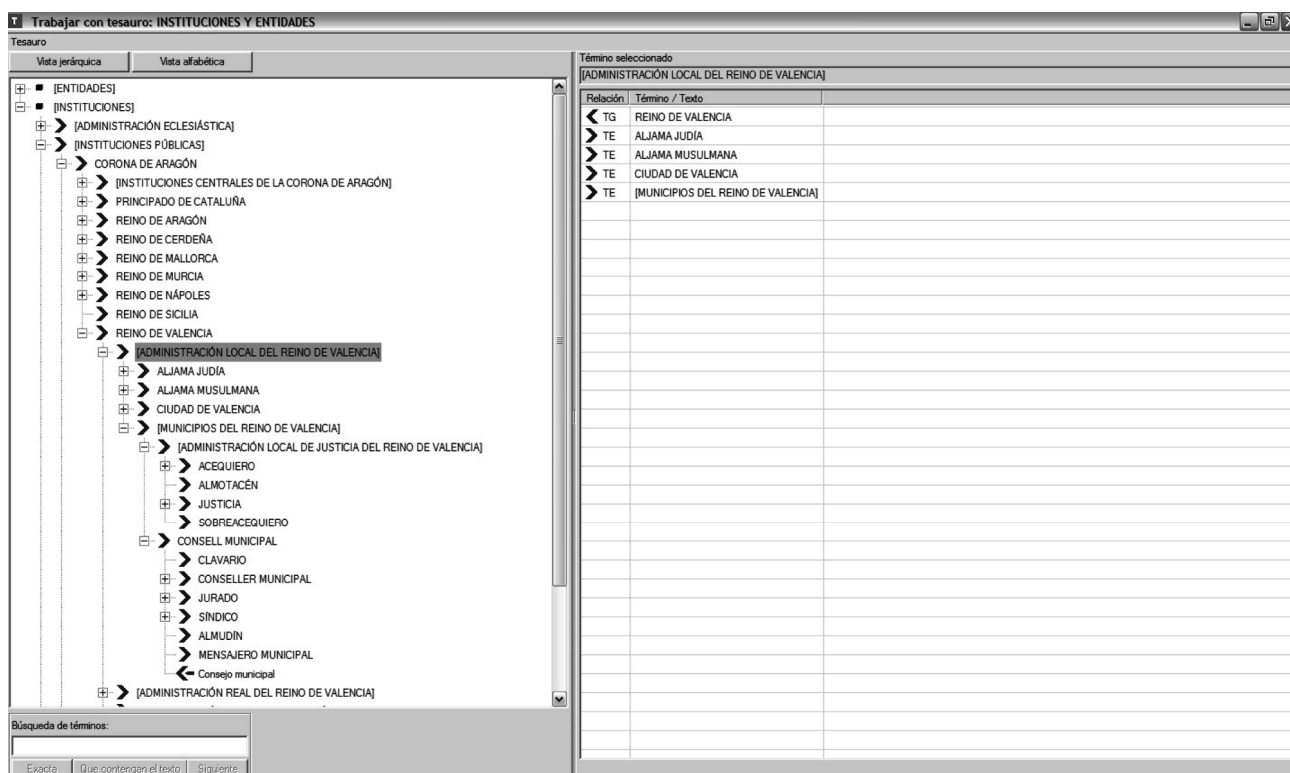

Figure 1. The hierarchical structure of the thesaurus for Institutions and Organizations, with ADMINISTRACIÓN DEL REINO DE VALENCIA highlighted. Source: Archivo del Reino de Valencia, Valencia 2006.

Knowl. Org. 38(2011)No.2
V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

163

"MESTRE RACIONAL DEL REGNE DE VA-LENCIA" is an institution that does not exist in the Kingdom of Castile. It therefore appears in Spanish as "MAESTRE RACIONAL DEL REINO DE VALENCIA," and, in the scope note, it is explained that "The maestre racional, established in 1419, is unique to the Kingdom of Valencia."

6.2 Semantic relationships with geographical descriptors

Mountains, rivers, and lakes have, throughout their history, been subject to name changes. The simple solution was to select the current name, but indicate the former names with terms that are not descriptors, redirecting the user to the descriptor. A similar solution could be applied when the same geographical entity is given different names in different places or regions. A slightly more complex problem was that of place names that, at one time, were local entities, but which no longer exist. Our solution was to construct a semantic relationship with those place names (states, provinces, towns, etc.) according to the legal situation at the time when the thesaurus was being set up. Former and now depopulated places were then incorporated as sub-terms under the name of the place which is responsible for it now. These sub-terms are descrip-

tors and are used as such when indexing documents belonging to the time when the entity existed. This involves research work and the need to consult bibliographical sources (encyclopedias, local histories, etc.) which may slow up the creation of the thesaurus, but which, once done, make the thesaurus far more useful.

In our region, for example, the town which is currently called GANDIA is a place name that has, throughout its history, absorbed other towns that have now disappeared, or have no place names of their own as they now form part of GANDIA. All of these towns, however, formerly had legal status as towns. Thus, the towns of BAIREN, ALCODAR and RAFACAN etc. were formerly towns in their own right, but which have now become depopulated. BENIOPA is a new town that, in 1965, was annexed to GANDIA, and therefore lost its legal status as a town in itself, as illustrated in Figure 2.

6.3 Semantic relationships with subject matters

The main problem with regard to semantic relationships was in placing a term within a hierarchy so that it could be used for more than one different concept. The best solution was provided by the information system with which we were constructing the thesau-
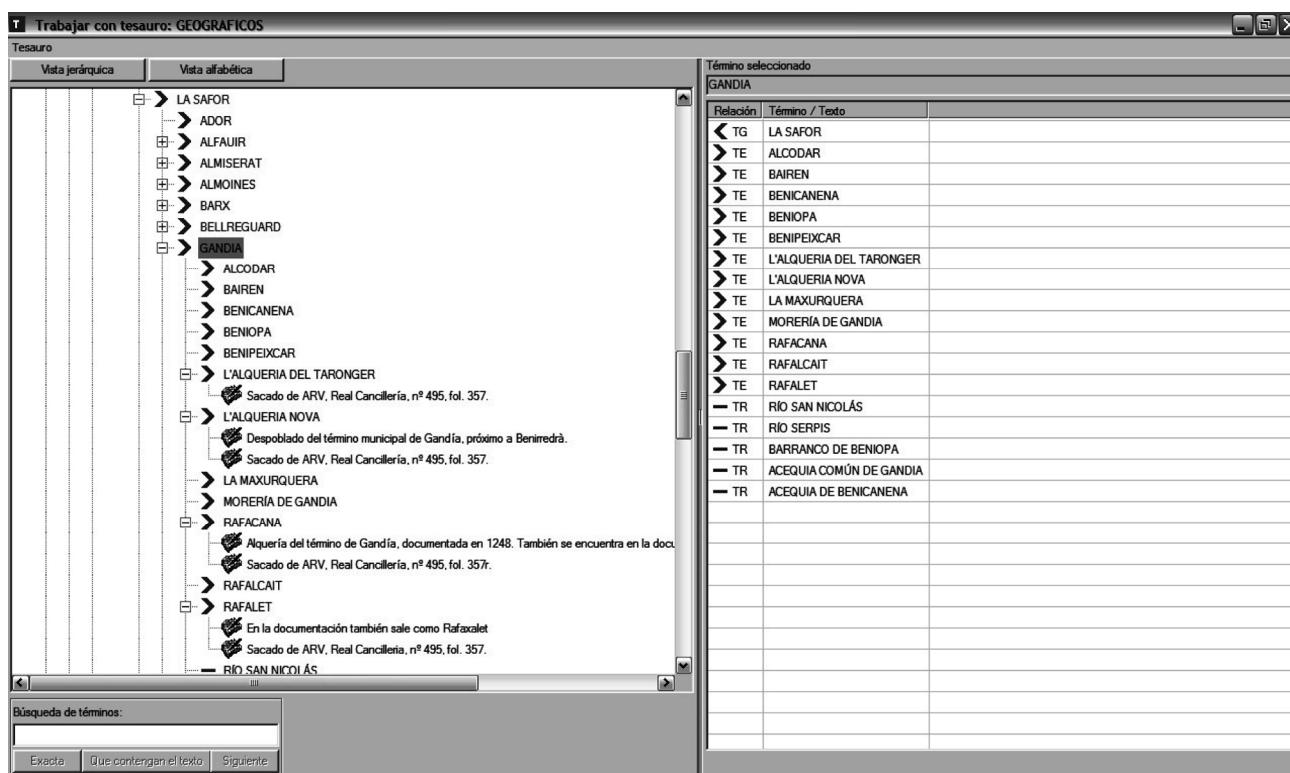

Figure 2. Hierarchical structure of the geographical thesaurus, with GANDIA highlighted. Source: Archivo del Reino de Valencia, Valencia 2006.

164   Knowl. Org. 38(2011)No.2

V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

rus, as it enabled the creation of multiple hierarchies. Although the descriptor itself remains the same, various different hierarchical relationships can be visualized under different headings. If the terms had undergone a slight semantic change in the course of time, a simple clarification was inserted. If the semantic change was more significant, however, the meaning in the predominant chronological period of the archive collection was chosen when deciding where the semantic relationship should be.

The concept of marriage (MATRIMONIO), for example, is now used with a wider range of meanings and uses than it was formerly, due to the separation of church and state in present day western culture. It was therefore necessary to use multi-level hierarchies in the thesaurus in order to indicate, in their corresponding hierarchies, its uses in the civil law branch as family law (DERECHO DE FAMILIA), or in the canon law branch, under sacraments (SACRAMENTOS), as illustrated in Figure 3.

7.0 Conclusions and plans for future developments

The reliability and functionality of the software used enables the four geographically separate archival cen-

ters of Valencia, Alicante, Castellón de la Plana, and Orihuela, along with the 13 archival specialists mentioned above, to use one server to create standardized descriptions using ISAD (G) and ISAAR-CPF, and to carry out thesaurus indexing. The database is uniquely dedicated to the three thesauri (for institutions, geographical place names and subjects) and is used to describe four classification groups, as each of the four archival centers mentioned above created its own classification group.

The possibility of offering one thesaurus for the four centers, and therefore to use one term for searching the documents in the four classification groups, has increased user interest (at the time of writing, mainly archivists) in carrying out document searches before physically visiting the city where the archive is located. Before the existence of the thesaurus, published inventories or index cards had to be consulted for each archive, which made searching all four of them simultaneously impossible. The users have therefore received the database with enthusiasm.

Although the issues involved in constructing an in-house thesaurus are not new, we have seen that very few previous examples are specifically related to historical archives. What has been described here is a
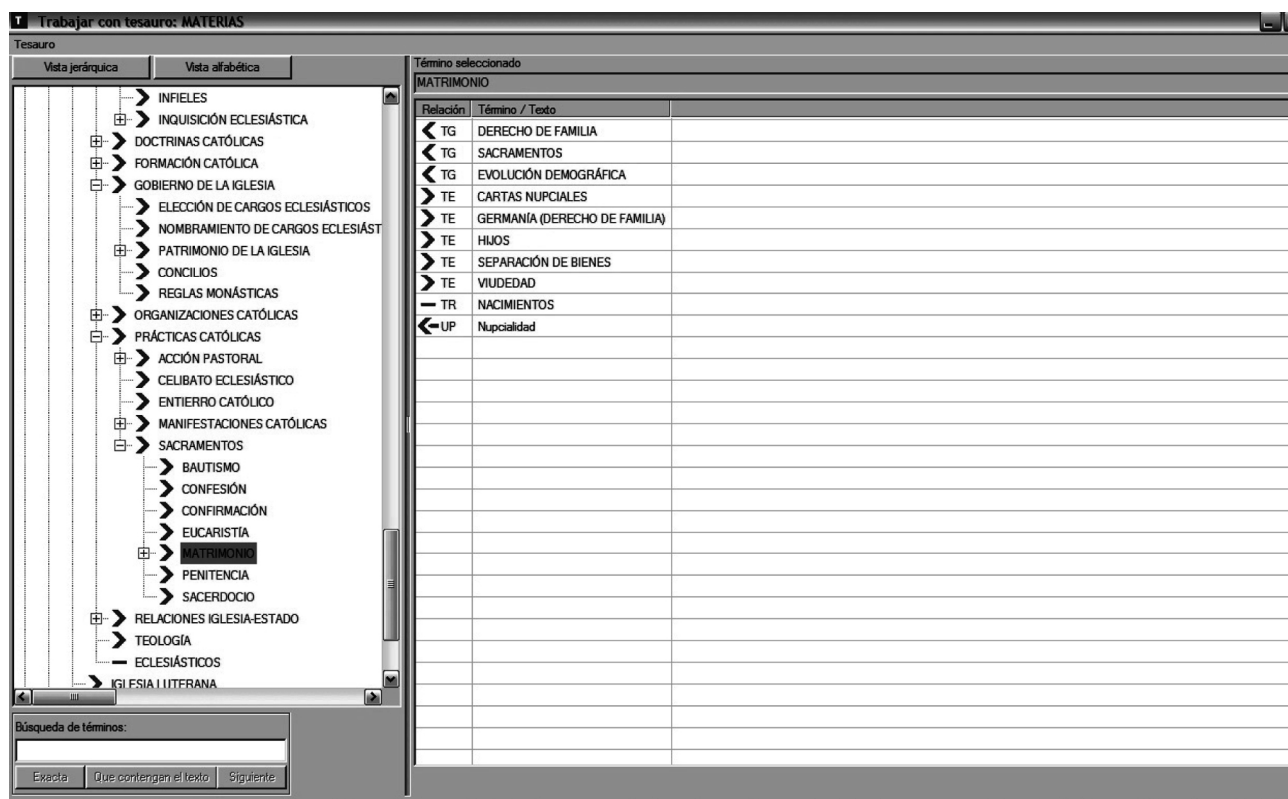

Figure 3. Hierarchical structure of the thesaurus for subjects, with MATRIMONIO highlighted, and showing the corresponding multiple hierarchies. Source: Archivo del Reino de Valencia, Valencia 2006.

prototype of a thesaurus incorporated into an automated and searchable system, which was the primary objective of creating a thesaurus in an intranet network for our four archival centers. With the technological and methodological issues now resolved, the thesaurus for these archives is now functioning on our intranet, and our project is now focused on making this tool available on the Internet. This will enable us to analyze the end users' experiences of information retrieval using the system devised and determine if the needs of researchers are being met. This, in turn, will facilitate research and, by extension, contribute to the sum of knowledge.

References

AIM25. 2002. *Archives in London and the M25 area*. Available http://www.aim25.ac.uk/search/index.stm.

Arano, Silvia. 2005. Los tesauros y las ontologías en la biblioteconomía y la documentación. *Hipertext.net* 3. Available http://eprints.rclis.org/8746/2/Los_tesauros_y_las_ontolog%C3%ADas_en_la_Biblio teconom%C3%ADa_y_la_Documentaci%C3%B3n.pdf.

Archives de Cannes. 2009. *Thesaurus historique.* Available http://archivescannes.ville-cannes.fr/4DCGI/Web_ThesConsult/listetreehisto/ILUMP4774.

Author?. 2002.Tesauro Archivo de la ciudad Arganda del Rey. Available http://archivo.ayto-arganda.es/documentos/TESAUROARGANDA.PDF.

British Library. 2009. *Integrated archives and manuscripts system (IAMS) cataloguing guidelines. part 2. Describing people, places, subjects and their relationships.* Version 1.0, pp. 1-13 [Unpublished report].

Cadieu, Céline. 1997. *Archives Modernes et contemporaines 1790-1997- État general des versements*, 2ᵉ ed.-État provisoire. Available http://www.archives-lyon.fr/static/archives/contenu/old/fonds/etat/vers.htm.

CIVIS. 2008. *Base de datos del sistema de Información municipal.* Available http//:www.civis.gva.es/civis/va/index_menu.htm.

Currás, Emilia. 1998. *Tesauros-manual de construcción y uso*. Madrid: Emilia Currás.

Direction des Archives de France. 2008a. *Thesaurus W, vocabulaire normalisé pour la description et l'indexation des archives administratives locales contemporaines.* Available http//: www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/thesaurus/.

Direction des Archives de France.2008b. *Indexation-état de la question.* Available http//: www.archives defrance.culture.gouv.fr/static/1083.

Direction des Archives de France. 2009. *Thésaurus pour la description et l'indexation des archives locales anciennes, modernes et contemporaines.* 4e ed. Available http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/thesaurus/.

Escrig-Giménez, Mercedes & Giménez-Chornet, Vicente. 2002. La informatització de l'Arxiu del Regne de València. In 2ⁿᵈ Jornades de l'Associació d'Arxivers Valencians. Torrent (Spain), 11-12 December 2002. [Presentation]. Available http://eprints.rclis.org/11468/1/experiencia03.pdf.

Escrig-Giménez, Mercedes, Giménez-Chornet, Vicente, Villalmanzo-Cameno, Jesús, & Viciano-Navarro, Pau. 2006. *Tesauros del Archivo del Reino de Valencia= Tesaurus de l'Arxiu del regne de València* [on CD-R]. Valencia: Generalitat Valenciana.

Gilchrist, Alan. 2003. Thesauri, taxonomies and ontologies; an etymological note. *Journal of documentation* 59: 7-18.

Gonzalez, Pedro. 1998. *Computerization of the Archivo General de Indias: strategies and results.*, Washington, DC: Council on Library and Information Resources. Available http//: www.clir.org/pubs/reports/gonzalez/contents.html.

Katifori, Akrivi, Torou, Elena, Vassilakis, Costas, & Halatsis, Constantin. 2008. Supporting Research in Historical Archives: Historical Information Visualization and Modeling Requirements. In Information Visualisation: proceedings of the 12ᵗʰ International Conference 8-11 July 2008 London, UK; IEEE Computer Society, Washington, DC, USA, pp. 32-37.

Informática el Corte Inglés. 2008. Archidoc. Available http//: www.iecisa.com.ar/archidoc.htm.

Instituto Andaluz del Patrimonio Histórico. 2008. *Tesauro de patrimonio histórico Andaluz.* Available http//:www.juntadeandalucia.es/cultura/iaph/tesauro/init.htm.

Mazzochi, Fulvio, Tiberi, Melissa, De Santis, Barbara, and Plini, Paolo. 2007. Relational semantics in thesauri: some remarks at theoretical and practical levels. *Knowledge organization* 34: 197-214.

Merlos-Romero, Magdalena, & Villarreal-Mascaraque, Eugenio. 2001. El tesauro de Archivos Municipales - Método y experiencias del Grupo de Archiveros municipales de Madrid. *Lligall* 17: 151-64.

Nielsen, Marianne-Lykke. 1999. User Associations: Do they have any value in a Corporate Thesaurus?

166   Knowl. Org. 38(2011)No.2

V. Giménez-Chornet and M. Escrig-Giménez. Designing a Thesaurus to Give Visibility to the Historical Archives ...

Proceedings of the 4th Conference ISKO-España. Granada, April, 22-24, pp. 117-125.

Péllisier, Jean.Pierre, & Motte, Claude. 2003 *Géonomenclature historique des lieus habités*. Available http//: www.archivesdefrance.culture.gouv.fr/static/1079.

Rubio-Liniers, María-Cruz. 1999. *Tesauro de historia contemporánea de España*. Consejo Superior de Investigaciones Científicas (CSIC): Madrid.

Sánchez-Alonso, Salvador, Sicilia-Urbán, Miguel Angel, & De Rato-Leguina, Gonzalo. 2008. Sobre la interoperabilidad semántica en las descripciones archivísiticas digitales. *Revista Española de Documentación Científica* 31n1: 11-38.

Santos-Lorite, María-Rosa. 2002. Un vocabulario para la indización del mundo profesional del Antiguo Régimen. *Scire: Representación y Organización del Conocimiento* 8n1: 131-145.

Sargent, Dick.1998. Introduction. In Gillman, Peter ed., *National name authority rile: Report to the National Council on Archives*. British Library Research and Innovation Report: Boston Spa (UK). Available http://www.nca.org.uk/materials/nationalname authorityfile.

Schenkolewski-Kroll, Silvia, & Tractinski, Assaf. 2006. Archival description, Information Retrieval, and the Construction of Tesauri in Israeli Archives. *Archival Science* 6:69-107, Available http://www.archives.org.il/UserFiles/File/silviaassaf.pdf.

United Kingdom Archival Thesaurus (UKAT). 2003. Available http:// www.ukat.org.uk.

UNESCO. 1995. *Unesco thesaurus: A structured list of descriptors for indexing and retrieving literature in the fields of education, science, social and human science, culture, communication and information*, 2nd ed. Paris: UNESCO.