

## **Subject determination during the cataloging process: an intensive study of five catalogers**

**Alenka Sauperl**

with

**Jerry D. Saye**

School of Information and Library Science  
University of North Carolina at Chapel Hill  
Chapel Hill, North Carolina

### **Abstract**

Subject headings and call numbers for the subject description of documents in library catalogs are provided by catalogers, who bridge the gap between producers (e.g., authors) and users of documents by developing representations to support information retrieval. While catalogers prepare document representations according to a set of standards and requirements and using subject heading lists and classification schedules, they also are in a position to select headings and classes that bring together authors' and readers' interests. In spite of the long tradition of subject cataloging, it is still not completely understood how this process works. The question that the research reported here investigates is how do catalogers decide about the topic of the document and appropriate subject description? Five experienced catalogers were observed and interviewed about their work and experience with subject cataloging in libraries. The think-aloud method was used for the observations, and unstructured interview was used for the follow-up discussion. A follow-up discussion clarified any uncertainties and explored some issues in more detail. The results were compared to the results of other similar research. The observed catalogers consistently performed subject cataloging in five stages: (1) identification of the topic of the book; (2) identification of the authors' intent; (3) inference of the possible uses; (4) relation of the topic to the existing collection; and (5) relation of the topic to the classification scheme and subject headings list. The sequence of stages was not necessarily linear, but was flexible in such a way that the catalogers returned to any of the previous stages whenever they realized they were on the wrong track. Their common strategy was searching for existing patterns in sets of call numbers and subject headings that had been assigned to the items in the existing collection. This study, limited in sample of five catalogers and thirteen books, does not allow for generalization over all catalogers. It is, however, useful in comparison with other studies and valuable in presenting different research questions that are still open in subject cataloging.

### **Introduction**

The main purpose of cataloging is to help people find information about library materials through library catalogs. Subject cataloging, a part of the cataloging process, focuses on providing information on the topic of the document, or on what the document is about. Subject representation is a critical element in providing access to materials. After all, the best information

retrieval mechanisms cannot perform well if the document representations are inadequate. That is, inappropriate assignment of subject headings can lead to retrieval failure -- the inability of the system to retrieve a relevant document.

A common complaint with subject document representations is the mismatch between terms used by searchers and the terms used in the information retrieval system (e.g., on-line catalog). It would be easy to blame inadequate subject headings lists (e.g., Library of Congress Subject Headings or LCSH) and classification systems (e.g., Library of Congress Classification (LCC) or Dewey Decimal Classification (DDC)) for this problem. Although this may be the case, subject headings and class numbers for subject description of documents in catalogs are provided by catalogers, who are the one mechanism that bridges the gap between producers (e.g., authors) and users of documents by developing representations to support information retrieval. While trained catalogers prepare document representations according to a set of standards and requirements and using subject heading lists and classification schedules, they also are in a position to select headings and classes that bring together authors' and readers' interests. In spite of the long tradition of subject cataloging, it is still not exactly understood how this process occurs, and the question that this research investigates is, thus, how do catalogers determine the topic of the document and appropriate subject description?

Although several textbooks attempt to present the process of subject cataloging, little research has been done on the process in a real environment. Jeng (1996) discussed knowledge acquisition methods for studying cataloging expertise. Those methods were used in her project in the Library of Congress where she was interested in the knowledge and skills of expert catalogers, their strategies for problem solving in cataloging, and specifics of cataloging instruction. She did not present the results of the project in her paper, but rather illustrated her research methods with two examples of the cataloging process. Her two participants cataloged (descriptive and subject) different documents (one copy-cataloging the other galley-proofs of a book to develop the Cataloging in Publication record) and exhibited different working processes. A question arises as to why the processes were different and whether Jeng's findings helped to explain those differences.

It was decided to look at the processes of subject indexing and abstracting, which can be seen as a process similar to subject cataloging. Indexing and abstracting occur in different circumstances and have different requirements than does subject cataloging, however, reading and the text summarization or extraction may be similar in all processes.

In one study on the process of indexing, David and her colleagues (1995) observed indexers at work and developed a problem-solving model of indexing. The model proposed a sequence of actions and required knowledge for assigning index terms to documents for a bibliographic database. It is assumed that the process of subject cataloging exhibits similar characteristics as indexing.

---

Although class number is only a part of call number, participants of this study usually referred to the call number meaning the class number. This synonymous use is followed throughout the paper.

Endres-Niggemeyer, Maier, and Sigel (1995) presented a detailed model of the abstracting process based on observing abstractors at work. The focus of the study was the development of meaning and the condensation of documents into abstracts. In addition to presenting a model with a sequence of actions as well as rules governing those actions in text reduction, they also described abstractors' strategies of reading documents and identifying the document's characteristics.

Textbooks on subject cataloging and manuals for the use of classification systems and subject headings lists usually suggest that catalogers read certain parts of the document to gain an understanding of subject content. It is not clear how those rules are observed in the actual working environment.

While some researchers observed indexers and abstractors at work, others contributed theoretical considerations to the processes. Beghtol (1986) considered the linguistic characteristics of communication and proposed a general model of classification and indexing. Her model suggested that every document can have a different meaning depending on the perspective of the author, his or her discipline, the reader, and the reader's current interest. Additional complications arise when the reader is the cataloger and develops the secondary text (subject headings). Afterwards this secondary text may have different meanings as well. Using this model one wonders whether the different meanings (discourses) play any role in the development of subject description.

Farrow (1994) discussed the indexing process on the basis of Kintsch and Van Dijk's model of text comprehension and reduction. Farrow stressed that all the developed models assume indexers have complete knowledge. Yet, more likely, indexers make decisions about index terms on the basis of incomplete knowledge of the document and of the discipline. Similarly Pinto Molina (1995) proposed that different disciplines apply different rhetoric structures to the documents. This means that each discipline has a "prescribed" document structure (e.g., sequence of chapters: introduction, method, results, discussion). That structure also defines what needs to be explained about the topic discussed in a particular paper and what can be assumed that a reader will already know but not included in the text. Some issues arise when Farrow's and Pinto Molina's ideas are used in the area of subject cataloging. Because of time constraints in cataloging departments it is very likely that documents are only partially inspected for subject content. In addition, catalogers in some types of libraries are probably not experts in every discipline in which they catalog. How do subject catalogers compensate for incomplete knowledge in their everyday routines?

Understanding the process of subject cataloging is important for several reasons. It can provide suggestions for improving subject searching in existing on-line catalogs. It may help editors of classification systems and subject headings to organize these tools in such a way that they would be easier to use. Further it might also aid in the training and education of new catalogers, which could ultimately result in better document subject descriptions in on-line catalogs.

In order to achieve these purposes, the goal of this current study of the cataloging process was to reveal some of that expertise of subject cataloging. Experienced catalogers were observed and interviewed about their work and experience with subject cataloging in libraries.

## Methodology

How do catalogers identify the subject of a document? How do they determine the appropriate call number and subject headings? To answer these questions, an exploratory study of the subject cataloging process in five public libraries in North Carolina was undertaken using verbal protocol analysis.

The think-aloud method was used for gathering data about the cognitive process of subject cataloging. With this method, participants are invited to say their thoughts aloud while they perform the observed task (van Someren, Barnard & Sandberg, 1994). The researcher is in the background during the observation, tape-recording and taking notes, only prompting the participant to proceed if necessary. This method provides a mechanism for learning about the participant's thought processes while minimizing the researcher's influence on the observed process (Meyer, Mniszewski & Peaslee, 1989). This main data gathering approach was supplemented by a short questionnaire on the cataloger's background and experience and a discussion following the observation of the cataloger's work.

In order to be included in the study each participant needed to hold a master's degree in library science, had worked as a cataloger in that library for at least one year, and perform original cataloging. The purpose of the study was explained to the participants as an interest in the cataloger's individual approach to subject cataloging. The observations of the cataloging tasks were scheduled during regular workdays.

Each cataloger was asked to create original catalog records for three different items which they selected employing the following criteria if at all possible:

- < book format;
- < for permanent addition to the general collection;
- < for adult readers;
- < selected from material that would normally be processed on the day of observation.

Catalogers were further asked to operate within the real environment of their cataloging systems and use any tools they wished. It was anticipated that the catalogers would not be able to perform the task automatically without paying attention and thinking about possible solutions, whereas with very simple examples they might perform the task automatically, routinely, which would be of little use to the study. Catalogers were asked to read aloud all the parts of the text in the books or on the screens they looked at, and to say aloud everything that came to mind.

A follow-up discussion in a second meeting included a non-structured interview to clarify any uncertainties the researcher might have about the first session and to gather information on the participants' feelings about their own work and contributions to the library. This discussion also gave the catalogers an opportunity to express their own thoughts about their work and seek the researcher's reaction.

Five catalogers from five different public libraries participated in the study. At the time of the study they had worked in their present positions for three to thirty-seven years. Their overall library work experience ranged from eighteen to thirty-seven years. Three of the five

participants were female, four had a bachelor's degree in English and one had a degree in music. They had received their master's degrees in library science between 1959 and 1983. Three of these master's degrees were from the same school but that professional education had been received in different decades (1950s, 1960s, 1970s). Two of these three participants had received their cataloging instruction class from the same instructor.

**Sample and method concerns**

Potentially there are some problems with observations using the think-aloud method. Although the participants received identical instructions to say aloud any text they looked at and to say aloud all their thoughts during the observation, sought the researcher's response to their activities and thoughts at some point during the session. Thus these observations were occasionally combined with elements of a non-structured interview. Elements of interactive-relational interviewing were used in this research; the researcher and participants interacted and shared a genuine interest in cataloging (Chirban, 1996).

In sum, the following actions were used to overcome or reduce the above problems:

- < explaining honestly the intentions and purpose of the study to respondents;
- < establishing an open and friendly relationship to gain trust;
- < projecting a calm and open attitude during observations and showing interest, appreciation, and respect without judging the work process; and,
- < triangulation - observation and discussion of respondents' work combined with a comparison with other researchers' work.

Class number	Class description	Collection		
		General	Reference	Local history
025	Operations of libraries, archives, information centers	1		
248	Christian experience, practice, life	1		
306	Culture and institutions (social sciences)			1
330	Economics		1	
362	Social welfare problems and services			1
371	School organization and management			1
629	Other branches of engineering			1
640	Home economics and family living	1		
745	Decorative arts	1		
929	Genealogy	2		1

**Table 1:** Books selected for observation by classes of the Dewey Decimal Classification

Another issue in the conduct of the study was the selection of books for the think-aloud sessions. It was necessary to modify the normal work flow to fit the research needs: books were selected specially for the observation; books acquired for a special collection were included; and two instead of three books were cataloged in two cases. Only two of the total thirteen books used in the observations were not published by a local publisher or governmental body. Catalogers commented that the policy of their libraries was to do copy-cataloging whenever possible with original cataloging discouraged and seldom performed.

Table 1 presents the subject characteristics of documents used in the study. The subject of the books was disregarded in this study and no consideration was given to the possible differences in the cataloging process for different subjects, although this may have influenced the results.

## Results

### Sequence of the cataloging procedure

#### The "mystery game"

"I consider cataloging a mystery game, like a detective story [pause] I feel like I've been given some clues about the book and I have to go to find the final clue \*\*"

This is a statement made by one of the catalogers, which captures the nature of the cataloging "game" as seen by this respondent. It expresses all the catalogers' willingness to accept the challenge and pursue the goal of providing access to the information. It also expresses their positive attitude, joy of doing cataloging, and curious, exploratory spirit. Accepting challenge means to search for the missing information in the interplay of all the available sources: the book, local catalog, reference books, colleagues and other sources. To reach the goal of providing the appropriate access to the information means that catalogers will never hear about the particular item again: there will be no complaint about an inappropriate call number or some other issue but also no prize. It also means that catalogers described all the aspects of the subject of the book that they were able to identify on their own and understood from the author. They also accounted for all potential uses they could anticipate within the framework of local practices and national standards.

The five catalogers consistently performed subject cataloging in five stages:

- < identifying the topic of the book;
- < finding the author's intention;
- < inferring possible uses;
- < relating the topic to the local collection; and
- < relating the topic to the classification scheme or validation.

The sequence of these stages was flexible so that the catalogers returned to any of the previous stages whenever they realized they were on the wrong track. Thus the stages interlinked so that they merged into one another gradually rather than suddenly.

---

\*\* Notation system for verbal protocols was adopted from DuBois, Schuetze-Coburn & Paolino (1993)

### The meaning of the book and the sense of "closure"

Catalogers first attempted to determine what the book was about. They all consulted parts of the books in the following order of increased informational content whenever necessary: title page, all parts of the cover or dustjacket, table of contents, introduction, index, and chapter titles. If all else failed, they read the first pages of selected chapters or entire chapters. They stopped when they felt able to describe the subject of the book in sufficient detail. The attainment of "sufficient detail" came from feeling of closure without passing the point of diminishing return. Thus if they felt confident about the topic of the book by only looking at the title page they only verified their decision by looking at the covers. If the information on the front and/or back cover confirmed the decision, they stopped. If not, they proceeded to acquire more information. One of the catalogers said: "I'm not getting anything new here [referring to the next part of the book]" in such a moment of closure. In cases where they were uncertain about the topic, these iterations continued until enough information was gathered to reach and confirm a decision about a topic or until no new information was gained.

An example follows from a cataloger going back and forth trying to accurately understand the contents of the book. The cataloger read the title "History of the hospitals in [geographical location]" a few moments before this example begins.

"... I will do first checking the text. I think first I will check the index and see if one or more hospitals is given a large amount of attention page-number-wise [looks at the index on the back of the book]. From the index we see the three hospitals are given fairly extensive treatment [reads names, thinks a few seconds] of course there are others and two are currently in existence [pause] since this is a local collection I think I may use the general heading and give individual subject headings to these two hospitals with extensive coverage [pause] and notice the subtitle says [returns to read to the title page] 'including sanitariums, infirmaries, and institutes ...'"

She had already determined one subject heading and was at that moment proceeding to consult the index for a more specific description. While doing that, she remembered the subtitle and returned to re-read it. Because she was not certain of the meaning of the concept, she looked for an explanation in the book and found it in a chapter title. A moment later, after some more thinking aloud, the cataloger decided:

"I will choose first 'hospitals--[geographical location]--history' like I say since these two hospitals receive fairly extensive individual coverage I will make an entry for each of those . and I'm going to check I've eliminated infirmaries and I'm not sure what an institute is [smiles and looks at the book again, flipping pages, pause] I'm going to check the book and see if I see an institute a specific institute listed [reads the title of the chapter] 'Institute for the curing of the liquor and opium habits' . so I guess that's not exactly . I was thinking that maybe a mental hospital but that is . it seems to be for alcoholic and drug rehabilitation."

In addition to transitions from title page, to index, to other parts of the book, this example also illustrates the effort to correctly understand the meaning of the concepts. Correct understanding

relates to appropriate identification of the author's intent, and translation of the concept (if necessary) into subject headings.

### **The search for the author's meaning**

The search for an author's meaning follows. Here the focus of concern is to properly represent the author's meaning. This feeling of representing and being responsible for the author's meaning has two dimensions. That is responsibility is understanding correctly what the author has written, and understanding what the author wanted to accomplish: in one cataloger's words, "why did he or she write the book?"

Understanding correctly what the author has written is closely related to identifying the topic of the book illustrated in the previous example when the cataloger wanted to understand the meaning of "institute." In that case the cataloger had to translate the term "institute for the curing of the liquor and opium habits," into the more contemporary term "alcoholic and drug rehabilitation."

This point is also illustrated in the comments from the follow-up discussion with another cataloger. She expressed concern about using appropriate terminology for transferring information from an expert (author) to a lay person (herself or a reader), to convey the information to the user "in everyday language."

"I hope that author has written either a preface or an introduction . hopefully these are only one page or two that summarizes the purpose of the entire document that gives the gist of what the author hopes to convey what they um had put so much effort into and hopefully the author has written this preface or introduction in everyday language that will give me some um additional clues as to the focus of the book [pause] because the text of the book itself might get somewhat technical or written very much for the professional in the field and not lend itself to a lay person [implying that she does not consider herself expert in every discipline for which she has to catalog books, pause] this is of course an education document [referring to one of the books cataloged in the observation session] there's a lot of terms that I'm sure mean a lot to educators . but for someone who's not a teacher or schoolboard superintendent um . for someone who who for example might be a parent have a child that attends one of these magnet schools um they're going to loose some in the jargon in the slang um just like there's a lot of jargon and slang in librarianship and when we have to start to explain things to nonlibrarians"

This comment also illustrates understanding what the author wanted to accomplish. It is linked to the inference of possible use to be made of the book. Catalogers were aware that the topic of the book could be interpreted from different perspectives and felt obligated to convey an author's original meaning first. Their primary belief was that the topic they identified as the message of the book was the author's intention. These catalogers wanted to transfer the author's meaning without distortion. This did not, by any means, imply that the author's was the only viewpoint catalogers wanted to convey.

### **The inference of possible users and uses**

In addition to seeking the author's intent, the catalogers also attempted to anticipate the use possible readers' would have of the document. This latter use may be other than the intent of the author. One of the respondents said "consider what would an average intelligent person who is interested in what would they ask for," implying that there may be more questions answered from the book than she was considering at the moment.

The same cataloger also took into consideration the point of view of the author of the back-of-the-book index:

"so we have the '[name] county--[state]--tithables taxation and county [state] . tithes [pause] ok so tithes and [name] county genealogy and taxation are the most useful and I can use slaves--Afro-Americans and slave holders if I want to . whoever did the index thought it would be useful"

Although the catalogers tried to identify as many different perspectives as they could, they usually limited the number of subject headings they wanted to assign in accordance with the local practice. In identifying possible uses of the book and balancing them with restricted number of subject headings, the catalogers considered needs of their colleagues as well as of the general public coming into the library. An example summarizes the reasoning behind the inference of possible users and uses:

"...we try to give access in as many ways as possible students use it [referring to the library collection] and a lot of serious researchers use it . it may save somebody some time."

During the follow-up discussion the catalogers commented that they had personal experience with users and that they communicated regularly with their colleagues. One cataloger summarized it this way without being asked:

"often something could be here or there it depends a lot on who you are doing classification for . where your library has put this kind of books before . whether the cataloger thinks it really needs to go here or there."

Restrictions in assigning terms also came from LCSH, where some of the document aspects could not be expressed, as in this example from the observation session:

"a lot of times I will use the broader subject heading . um . but in this case I don't think it is appropriate I mean it would be misleading [pause] I'm looking at these other subject headings just to get an idea [reading from the Library of Congress Subject Headings] CD ROM books . no . CD ROM catalogs it says use CD ROMs dash catalogs . CD ROM industry now here for manufacture of compact discs use . that's not appropriate either . so it looks like CD ROMs is the only subject heading that we can use"

### **The relationship to the local collection**

“What call numbers and subject headings have we assigned to similar books in the past?” was a common question asked by catalogers when they began to translate the identified subject(s) of the book into a call number and subject headings. This question was resolved by searching in the local catalog. This search was by either subject or keyword depending on whether the catalogers were searching for a class number or subject headings. One cataloger described this:

“I go to the local system and see if we’ve got books with the same subject heading and what kind of call number do they tend to have [pause] and if there is a consistent pattern of those for example the first subject heading on the OCLC screen was home economics that’s pretty much going to put us into the 640 section [pause] if I find that most of the other books that we already own here in the library as indicated in the [local system] have these subject headings and a pattern of consistent call numbers that’s what I’ll go with first.”

Another cataloger performed a search and talked aloud:

“I’m doing subject keyword search and I’m typing in hospitals and history [looking at the screen and reading aloud only part of the retrieved record] no edition given at the number this number came from Dewey 19 let me check and see [verifies in DDC 21]”

The catalogers often returned to looking at different parts of the book (basically returned to the first stage of identification of the topic of the book), if the call number or subject headings from the catalog didn’t convey the same meaning that the catalogers wanted for the book. In this particular example, the cataloger had considered some subject headings before and performed a search in an authority file, but was not satisfied with the result. She returned to the book and the local catalog because she wanted to find a better, more informative subject heading. The following is from the think-aloud session:

“for the moment we have ‘love -- religious aspects -- Christianity’ seems the title indicates . if this was not a local publication I may left out this African-American . I feel I need to bring that in some way . and I don’t think that African-American -- [geographical location] would be appropriate . I’m going back to the subject authority file . and then see if I can assign any heading that would [stops talking] religion . I’m not sure maybe I want . no I don’t like this it’s not exactly what I want [pause] we’ve assigned the subject heading ‘Christian life’ to one of her [author’s] other titles . I’m going to try that and see if in any way I can relate that . let me go to our local system and try something here I’m going to try subject keyword search . I’m searching Christian and try to pull Americans . oh [disappointed] . I’m going to try again . directly . I’m going to try ‘American women’ [gets the results and reads subject headings] see what we get . intellectual life . religious life .let’s try now maybe ‘conduct of life’ . let me go back and see [returns to read the book] let me see if there is only African-American religious life [turning pages of the book] although it certainly goes beyond that...”

Searching for a subject or keyword in the on-line catalog typically resulted in a screen display of a list of items, consisting of author, part of the title, year and call number. This information, although fairly limited, was the basis for determining the similarity of the item it represented with the item that is being cataloged at the moment. The selection process consists of seeking patterns, an issue discussed later in this paper.

#### **Verification of the subject representation with the classification scheme**

Rarely did the catalogers feel absolutely certain that the chosen call number and subject headings were totally accurate. In spite of their long experience for most books they verified the chosen terms with DDC schedules or LCSH. This verification had two dimensions. The first was to verify if the chosen term was valid according to DDC or LCSH. Example of this dimension of verification comes from one of the previous examples:

“...no edition given at the number this number came from Dewey 19 let me check and see [verifies in DDC 21]...”

Another example illustrates the verification of term as a valid subject heading:

“... infirmaries needs to be cross-referenced under hospitals its not a valid heading and institutes I still try to find...”

The second dimension of verification was whether the cataloger had interpreted and understood the term correctly. One cataloger said:

“ok [reads DDC 21] 004.56 is ‘external auxiliary storage’ this edition kind of changes a little bit [from the previous DDC20, which was published in fall 1996, implying that DDC 21 was still fairly new] . but I don’t think that we are talking about here is the CD ROMs themselves so much as it is the information that’s stored on the CD ROMs so I don’t think 004 is where we want to put this book so I’m going to look at the 025 numbers that have been suggested [in the index of DDC schedules] and see [...] ok 025.344 is ‘machine readable materials class here CD ROMs computer software interactive video multimedia’ [reads from DDC21] [...] so probably we could toss a coin between 004 and 025 [ ... ] but this would better serve when put in the library and information science number rather than 004 because I think that in the 004 number that is if you look at it you know you go back to once again for the discipline [turns pages back to the main class] you’re going to have more information about how I would think that it would be more appropriate information about how CD ROMs are made for example . the efficacy of them as a storage medium um . more technical kind of information about CD ROMs I think would be the emphasis there rather than listing or a catalog of CD ROM products...”

Call number verification occurred immediately following selection of that number. If the chosen call number failed to express what the cataloger wanted he or she returned to the book or to the local catalog to begin the process again using different words. If verification showed that the chosen number was changed in a later edition of DDC, the cataloger altered it accordingly and

concluded the task. Verification of subject headings occurred in a similar way. A cataloger consulted subject headings in the authority file to decide which term to use in the particular case:

“...and we’re going to look OK so we have street cars and busses . I’m going into subject authority [stops talking, searches for] street cars street car yeah street railroad [long pause, reading the screen] its such a strange subject heading [not talking, searching] [finally reads from the screen] local transit . OK so they want us to use street railroads.”

### Strategy of patterns

The common strategy of the catalogers was to look for similar books in the local catalog or OCLC in one of two ways. One approach was to use existing records to create the new record. The existing record was used as a pattern and source for the new record. The older record’s data were included in a new workform and used as a template, which was changed for the new item during the cataloging process and finally saved as a new record in the system.

Catalogers chose a similar existing record in several ways with the idea of saving time. They wanted to avoid some difficult task. Having a pattern or format helped them make sure that they did not have to type it again or verify it separately with the authority file. They may have decided to choose a record with the same author. With that they avoided verification of the personal name form in the authority file and typing errors in the author’s name. In one instance, the cataloger chose an author who wrote on a particular subject. The existing record not only gave the authorized form of the personal name entry but also reasonable suggestions for call number and subject headings. Catalogers may have chosen records with other bibliographic elements that they consider important such as call numbers and subject headings. Another cataloger decided to choose a previous edition of the item. This use of patterns appeared during the whole cataloging task. In this study, subject cataloging was only one, inseparable part of the whole cataloging process. Thus, while attention here is on the subject cataloging process, it is interwoven with descriptive cataloging.

The second approach to using existing catalog records was to identify call numbers and subject headings that co-occurred in individual catalog records. Certain call numbers and subject headings or combinations of subject headings seem to appear together. These catalogers desired to follow these patterns in order to promote consistency. On occasions, when they wanted to assign the same combinations of call number and subject headings to the new item so they searched for appropriate records in the catalog. In the following example, one cataloger wanted to use a record he remembered from a few days before.

“... I’m going to do a subject search . and in crafts and it turns out that the other day I had another book that I was cataloging that was on the same subject that I did find information on and I learned that it was crafts and I believe it was management that I found [he searches for crafts and management] . no that wasn’t it [surprised with no matches (0 hits), pause] am perhaps it’s too broad [thinks quietly then says to the researcher] can you hold on a second when I go find the book.”

Just a little later he explained:

“I do a lot of stealing when I catalog . I look at other records that are similar to what I’m cataloging and you know if they are close enough I will go ahead and if appropriate use those subject headings and call number.”

All the catalogers had similar memories of previous examples. This particular example, however, also illustrates how human memory can fail and that the cataloger needed to find a way to compensate.

A cataloger referred to this “stealing” as “borrowing” and another called this behavior “chicken cataloging.” On other occasions, catalogers chose to follow the pattern for clarification, as in the case of one cataloger who wanted to assign the geographic subdivision of a subject heading for a case, where she was uncertain about its use and the LCSH instruction also was not clear.

The local collection played an important role in providing these patterns although OCLC was occasionally used in a similar. It appears that the main reason for following patterns in subject description from local catalogs was indexing consistency, whereas OCLC was used as a standard or source of ideas and recent practice. In the observed cases, OCLC records were adjusted for local practice, and local patterns were changed to follow examples of apparently new OCLC or LCSH guidelines.

#### **A quiet finale**

Restrictions and obstacles of the classification systems and local library practices are problems daily encountered by the catalogers. Additionally, these catalogers indicated a communication problem which complicated their confidence in balancing those restrictions and assigning suitable subject headings and call numbers. When catalogers did not get feedback from their colleagues they assumed that appropriate subject headings and call numbers had been assigned. One cataloger stated:

“ and if librarians who receive these books don’t protest or say wouldn’t this be put better somewhere else then you think that’s OK”

Not receiving any comments, negative or positive, meant that colleagues considered their cataloging well done. In a follow-up discussion one cataloger commented: “we are the invisible ones . only if we do something wrong they [referring to colleagues] notice us”.

During the observation session one cataloger talked of making a final check of the new record before saving it:

“...he likes the record now [referring to himself] . I’m always uncomfortable if I can’t find exact match for what am I creating . but yeah he’s happy [referring to himself again]...”

Such final verifications of records were routine. The catalogers verified their own work after they had considered the rainbow of aspects: the author, user, local catalog, WorldCat catalog, cataloging rules, subject analysis tools, and related resources or local practices. Each of these is important at least for a moment or two in the described stages. It was not possible to see, from the data available, that one or the other aspect dominates the record, influencing the local and OCLC catalog. It appears, however, that patterns play an extremely important role in catalogers' memory of previous examples despite the known fallacy of human memory -- forgetfulness.

## Relationship of the findings to other research

### Cataloging study

There are few published reports of research on the process of subject cataloging. Jeng (1996) presents verbal reports as a research method illustrating the method with two examples. One of them presents a protocol of the cataloging process (descriptive and subject cataloging) of one cataloger. This is an example of a cataloger in the Library of Congress, working on a cataloging-in-publication record for a galley proof of a book (preliminary pages, table of contents, and first chapter instead of the whole book). This cataloger browsed preliminary pages and the table of contents to identify the topic of the book. On this basis, he decided on some subject headings. After searching for the author's name in the local catalog and noticing a phrase appearing in the titles of subject headings of other items by the same author, he proceeded with the title search in the same catalog with that phrase and consulted subject headings assigned to the retrieved items. Because he was not satisfied with the headings chosen so far, he returned to the item at hand and considered parts that he did not read before. He decided to drop one heading and add two. He then searched the two headings in the authority file and local catalog and read notes in the authority file and the subject headings of the retrieved records. As a consequence of information from this search, he decided to subdivide one subject heading and add a second one. He repeated the search in the authority file and the local catalog for the new heading. With new information, he decided to change the assigned subject headings again. Looking at the whole subject description, he was satisfied and finished the subject description by assigning a classification number.

This description is partly in agreement with the sequence of procedures identified in the current study of cataloging process. Jeng's cataloger identified the topic of the book and related the topic to the local collection and to the classification scheme. He did not look for author's intent to verify his own perception of the book's topic. He was not interested in inferring the possible use for the book to add other aspects in the subject description. The circular pattern of moving among the book, the catalog, and the classification scheme is present in Jeng's example as well. It is only narrower, with three instead of five steps. Her participant developed the subject description he believed was the best by adding and dropping subject headings while acquiring information in the three sources.

It would be interesting to know why the users' needs did not play a role in Jeng's example. There seems to be no reason for this aspect to be unique to the five catalogers in this study on cataloging process. Did Jeng's cataloger ignore the author's intent or was that the only aspect he considered? It may be the case that considering the users' needs is stimulated by the immediate

presence of the users, which is less obvious in a very large library such as the Library of Congress. It may also be the case that author's intent appears as the counterpart to the users' perspective.

Jeng's cataloger sought inspiration from previous records as did the catalogers in this study. What role does following patterns or examples actually have? It may help with consistency of the catalog. It may also be a shortcut for the catalogers, when less verification in different files is necessary. Most of the time, it seems, the patterns are a kind of memory aid, helping catalogers to remember what else should be described, despite the possibility that one will forget or remember previous examples incorrectly.

### **Indexing study**

There is more research in the indexing area, which can be considered a similar process. Fugmann (1993), for instance, suggests that indexing is "the two-step process of a) discerning the essence of a document and b) representing this essence with sufficient degrees of predictability" (p. 67). Fugmann also states that the "classing of documents by means of classification schedules is included in this definition [because] classification notations are assigned to the documents instead of natural language terms" (Fugmann, 1993, p. 67). Because of this similarity and lack of the research on the subject cataloging process, it was decided to also compare the findings here with results of research into the indexing process.

David and her colleagues presented a model of the indexing process (David et al., 1994). They observed eight indexers during the indexing task and videotaped this session. Researchers discussed with indexers the videotaped session and also invited them to comment on other indexers' keywords produced for the same document that participants indexed for the study. Researchers took the perspective that indexing is a problem-solving task and divided the process into two general areas: knowledge space and resolution space. The knowledge space contained all the necessary knowledge for performing a task: indexing knowledge (procedures), vocabulary of the discipline, thesaurus (indexing language), and context knowledge (policies, users). Resolution space described the basic operations in the process: scanning of a document (recognition of the structure and aboutness of a document), content analysis (extraction of the essence of a document), concept selection (recognition of important and informative concepts), translation of the concepts into the vocabulary of the controlled vocabulary (assigning of index terms), and revision (evaluation and reconsideration of decisions made).

Indexing knowledge, including procedures, could be compared to the knowledge of cataloging rules and the cataloging system in this study of the cataloging process. Vocabulary of the discipline, which is necessary for the understanding of the topic discussed in the document being indexed, is part of the identification of the topic of the book and inference of the author's intent. Thesaurus or indexing language is parallel to DDC and LCSH. Context knowledge, which includes knowledge of policies and users, is included in the sequences of inferring possible use and relating the document to the local collection.

The sequence of procedures found for the catalogers in this study of the cataloging process, can be compared to the resolution space of the indexing model developed by David et al. (1994). Identification of the topic of the book is represented in David's indexing model in three separate

operations: (1) browsing of the document for recognition of the structure and aboutness of a document; (2) content analysis for actual extraction of the essence of a document; and (3) concept selection, meaning recognition of important and informative concepts. The two procedures of the indexing model, translation and revision, are parts of the two sequences of relating the item to the local collection and to the classification scheme. Translation of the concepts into the vocabulary of the controlled indexing language for assigning index terms is similar to the assigning of call numbers and subject headings, which occur throughout the sequence, but is mostly present in relating the identified topic to the local collection and to the classification scheme. A local collection serves as a source of ideas and a classification scheme as a standard for verification. Revision for evaluation and reconsideration of decisions is included in both, relating the identified topic to the local collection and to the classification scheme, because verification and changes occur in both stages of the sequence.

David's study identified most of the operations that were also identified in this study of the cataloging process. The consultation of similar records (strategy of patterns) is not evident, which was mentioned by Jeng's cataloger. It would be interesting to know if in fact only the issue of relating the new item to the existing collection in terms of seeking inspiration for subject headings (index terms) is different for indexers than for catalogers or if this strategy comes from a particular type of library environment.

### **General model of classification and indexing**

Does the image of the subject cataloging process developed in this study of five catalogers fit a general model accepted by the scholarly community? Beghtol's model of classification and indexing suggests that the reader's comprehension of the text is influenced by (1) the author's intention; (2) the reader's purpose; (3) the reader's culture; and, (4) discipline (Beghtol, 1986). This reader can be an indexer or someone else. In this case, the reader was the cataloger. Every reading and comprehension task is multi-layered in a situation of text summarization or classification. Those layers are: (1) text to be comprehended and expressed; (2) the classification scheme or indexing language; and, (3) the selected terms (call number or subject heading). Beghtol's model implies that catalogers need to consider all four comprehension influences when engaged in subject analysis and also all three task layers. The process of the catalogers in this study of the cataloging process compares with Beghtol's model as follows:

- < Identifying the topic of the book is the cataloger's understanding of the document of the general model. Participants in the study of the cataloging process were aware of possible subjective interpretation (reader's culture and discipline can be different from the cataloger's) and wanted to minimize it.
- < Finding the author's intention is one of the measures against subjective interpretation of the topic of the book. Beghtol's model describes this as the author's meaning of the text. Catalogers acknowledged that authors can have a different perspective on their books.
- < Inferring possible use is the other measure against subjective interpretation of the topic of the book. Catalogers acknowledged possible different perspectives that users could have on the book and wanted to provide ways that would enable users to find the book in the subject description.

- ⟨ Relating the topic to the local collection is parallel to the real environment in the general model. We assume that the catalog represents the current reality. Although this may not be entirely true, it is as close as we can come to the reality.
- ⟨ Relating the topic to the classification scheme is Beghtol's "tradition." The cataloging tradition can also include local practices, cataloging rules, standards for cooperating in the cataloging system, and principles of the cataloging activity. The observed catalogers respected this tradition as the only means that enables communication in the community of librarians and users.

Findings from the study with catalogers are in accordance with Beghtol's general model of classification and indexing. Clearly the catalogers attempted to balance different points of view in reading and comprehending a text, as well as to anticipate possible uses of the document. It is not clear if one of the perspectives is considered more important and when. It is also unclear what the implications are of the interplay of the perspectives for the consistency of the catalog and retrievability of the records. Catalogers may have different success in this effort. It was not measured, however, their awareness of the problem was identified

## CONCLUSIONS

In this study of the subject cataloging process five professional catalogers in public libraries were observed using the think-aloud method for the observations, and unstructured interview for the follow-up discussion. The observations provided data about the process. The follow-up discussion was designed to clarify any uncertainties and explore some issues in more detail. Both were audiotaped and transcribed. Verbal protocols were analyzed, and the results were compared to the results of similar research.

The sequence of procedures that emerged from the generalization of individual procedures of the five catalogers consisted of five basic steps: (1) identifying the topic of the book, (2) identifying the authors' intent, (3) inferring the possible uses, (4) relating the topic to the existing collection, and (5) relating the topic to the classification scheme and subject headings list. The catalogers' common strategy was searching for existing patterns in sets of call numbers and subject headings that had been assigned to items in the existing collection.

A basic skill for the identification of the topic of the document is reading. Detailed insight into what and how catalogers read, or what is sufficient detail, was not obtained in this study, but could be the subject of future research. Such a focus could provide potentially interesting implications for the education of catalogers. Suggestions for reading, its focus and detail, could be discussed in classes, which would at least provide an awareness among novices.

Complementary to reading is the structure of the document. Although not part of the study, the importance of this area is evident in other research as well as in the actions taken by the catalogers observed for this research, who emphasized certain parts of the documents over others in making their cataloging decisions. Graphic design and internal structure of the document (meaning, structure of the message) are both important for the catalogers' reading and understanding of the document. This suggests another area for future research.

The next stage in the sequence, identification of the author's intent, is quite controversial. Other research acknowledges the existence of the author's point of view. But it is difficult to evaluate whether the catalogers correctly understand what the author has written (if there can be one correct understanding at all). It can be expected that this problem would be bigger in public libraries, where catalogers are trained in one discipline and need to provide subject description for documents from a wide range of different disciplines. This can be less problematic in specialized indexing services, where indexers provide subject description for documents from the same discipline they were trained in. During training, new catalogers should become aware of the importance of general education and life-long learning.

A similar problem holds for the stage of inferring the possible uses of a book. It seems impossible to anticipate all the possible ways a book will be used in the future. Catalogers took a practical approach to this task by building on their knowledge about past and present users' needs as well as the needs of their colleagues. While the latter seems more realistic, assuming that the catalogers come in contact with their colleagues more frequently, the former seems more problematic. How accurate can an assumption about users' needs be if the cataloger does not meet users? If subject description is prepared according to users' needs in an individual library, one can doubt how universal it is and how useful it may be in OCLC and other libraries' catalogs (after copying). Nevertheless, two things seem crucial: ability of the cataloger to communicate with colleagues in the library and knowledge of user models and users' needs.

The stage of relating the book to the local collection is interesting because of its possible influence on indexing consistency in the local catalog. It could be expected that indexing is more consistent and consequently the catalog more useful if similar books are described with similar combinations of class numbers and subject headings. The problem is in establishing the characteristics for evaluation of this similarity. Considering the observed evaluation, in which catalogers looked at very limited information about the documents, this provides even more reason to doubt the accuracy of those evaluations. The efforts of catalogers to apply existing patterns of subject description seem to be aimed at promoting cataloging consistency. Understanding this process, however, could help in training novices by warning against exploiting this approach.

The stage of relating the item to the classification scheme or verification is the one that at least provides the use of the same expression through the different editions. It does not mean that we all interpret the same topic with the same class number or subject heading. We can probably agree that the classification scheme or subject heading list is a common language for communicating the topic of documents. A part of learning about the scheme might also be shared understanding of the concepts represented in the scheme.

This study, limited to a sample of catalogers and books, does not allow for generalization across all catalogers, yet it is valuable in presenting different research questions that are still open in subject cataloging. It is also useful in comparison to other studies: one study of the cataloging process, one of the indexing process, and a general model of classification and indexing. Although different in approach, they reveal the same concepts and similar procedures. It also is noteworthy that the catalogers' sources of inspiration in the selection of subject headings coincide with the sources identified by intermediary searchers in a recent study conducted by Iivonen and Sonnenwald (1998). This paper, published after this study of catalogers was

completed, reports that the intermediary searchers derived search terms from six discourses: controlled vocabulary, documents and the domain, indexing practice, users' search requests, existing database, and previous search experience. The coincidence of these discourses with the five cataloger's sources of inspiration for selection of subject headings seems to suggest similarities among people on the generating and using sides of the bibliographic databases.

### Acknowledgments

Our first thanks go to the five catalogers in the public libraries of North Carolina who participated in the study. Their kind and generous support made this study possible.

The authors also wish to express their appreciation to Dr. Paul Solomon of the School of Information and Library Science, University of North Carolina at Chapel Hill, for his advice on methodology during data gathering and analysis and for his helpful comments on earlier versions of this paper.

### References

- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42, 84 - 113.
- Chirban, J.T. (1996). *Interviewing in depth: The interactive - relational approach*. Thousand Oaks, CA: Sage Publications.
- David, C., Giroux, L., Bertrand - Gastaldy, S., Lanteigne, D. & Bertrand, A. (1995). A cognitive approach to consistency. In Kinney, T. (Ed.) *Proceedings of the 58th ASIS annual meeting*. (pp. 49-55). Medford, NJ: Information Today.
- DuBois, J.W., Schuetze-Coburn, S.C. & Paolino, D. (1993). Outline of discourse transcription. In Edwards, J.A. & Lampert, M.D. (Eds.), *Talking data* (pp. 45-89). Hillsdale, NJ.: Lawrence Erlbaum.
- Endres-Niggemeyer, B.; Maier, E. & Sigel, A. (1995). How to implement a naturalistic model of abstracting: Four core working steps of an expert abstractor. *Information Processing & Management*, 31, 631-674.
- Farrow, J.F. (1991). A cognitive process model of document indexing. *Journal of Documentation*, 47, 149-166.
- Fugmann, R. (1993). *Subject analysis and indexing*. Frankfurt: Index Verlag.

- Iivonen, M. & Sonnenwald, D.H. (1998). From translation to navigation of different discourses: A model of search term selection during the pre-online stage of the search process. *Journal of the American Society for Information Science*, 49, 312-326.
- Jeng, L.H. (1996). Using verbal reports to understand cataloging expertise: Two cases. *Library Resources and Technical Services*, 40, 343-358.
- Meyer, M.A., Mniszewski, S.M. & Peaslee, A.T.Jr. (1989). Using three minimally biasing elicitation techniques for knowledge acquisition. *Knowledge Acquisition*, 1, 59-71.
- Pinto Molina, M. (1995). Documentary abstracting: Toward a methodological model. *Journal of the American Society for Information Science*, 46, 225-234.
- van Someren, M.W., Barnard, Y.F. & Sandberg, J.A.C. (1994). *The think aloud method: A practical guide to modeling cognitive process*. London: Academic Press.