Guidelines for Subject Analysis in Subject Cataloging

Franciele Marques Redigolo, Mariângela Spotti Lopes Fujita & Isidoro Gil-Leiva

Published online: 04 Aug 2022.

Submit your article to this journal ⬚

Article views: 315

View related articles ⬚

View Crossmark data ⬚

Routledge
Taylor & Francis Group

Check for updates

Guidelines for Subject Analysis in Subject Cataloging

Franciele Marques Redigolo[a] (iD), Mariângela Spotti Lopes Fujita[b] (iD)
and Isidoro Gil-Leiva[c] (iD)

[a]Information Science Department, Federal University of Pará – UFPA, Belém, Brazil; [b]Information Science Department, São Paulo State University – UNESP, Marilia, Brazil; [c]Information Science Department, University of Murcia – UMU, Murcia, Spain

ABSTRACT

The representation of information in subject cataloging as a result of subject analysis will depend on the cataloger's prior knowledge, influenced by subjectivity. The subject analysis in cataloging is a central theme of this investigation with the aim to elaborate guidelines for subject analysis in cataloging. For this purpose, how books are cataloged in university libraries has been verified. The Individual Verbal Protocol was applied with catalogers from Brazilian and Spanish University Libraries. Directions for the elements and variables of the subject analysis and procedures for good development were obtained to constitute the Guidelines of Subject Analysis in Cataloging. It is concluded that the guidelines formed by four sections are indicated for incorporation in subject cataloging procedure manuals for the purpose of improving the levels of representation and information retrieval results.

Introduction

Subject analysis is a fundamental part of subject cataloging, and its practice depends on standards and methodologies due to its purpose of representing documentary content for future retrieval in subject search systems. However, this practice seems to be overcome by automatic extraction technologies for content representation that do not require human content analysis. Another revealing aspect is the lack of updating of subject analysis standards.

This study aims to develop a theoretical model for the creation of guidelines for accessing documentary subjects that apply to the cataloging of books in any library, of any culture, linguistic group, size of the collection, and users. However, defining variables for each context and a specific situation is necessary for this purpose. The elaboration of guidelines can be time-consuming and complex, nonetheless, once developed,

a cataloger's work in subject cataloging can advance, reducing the subjectivity of the process and increasing the cataloger's agility.

From the perspective of the existing norms on subject analysis, an analytical study of methodologies, and an observation study of the cataloging of books are proposed to elaborate guidelines for subject analysis in cataloging. For the observation study, the introspective technique of Individual Verbal Protocol was applied with catalogers of Brazilian and Spanish university libraries. Ethnographic research was used to obtain feedback from the cataloger on the process of representing the subject of books in subject cataloging. Brazilian and Spanish catalogers were selected as participants in this research because they work with books in standardized formats, whose textual structures are stereotypes known worldwide.[1] Familiarity with the textual structure is important so that the cataloger can rely on the text to perform the analysis more confidently.

Three important issues are raised by Taylor and Joudrey[2]: first, "in a profession that often seems to devalue the functions of technical services, it is important to recognize subject cataloging as an important skill for all aspects of librarianship"[3] and second, when "they disagreed with the tendency of closing cataloging departments and taking the jobs of professional librarians with the idea of saving money in large libraries."[4] And finally, when they wrote: "It can be tempting to underestimate the importance of subject heading searches, with the idea that searching for keywords will solve all our problems. This moment has not arrived yet (and may never arrive). In this "information society," we live in a complex world, where information retrieval is not just a walk to the catalog of records or a walk on the Internet [...]. Our world of information is crowded, confusing, and often frustrating. Access to subjects, controlled vocabulary and classification remain important and are highly underused."[5]

Taylor and Joudrey wrote on this subject in the early 2000s,[6] just before the "tsunami" of mass incorporation of e-books into libraries that dealt another near-fatal blow to cataloging and classification by library professionals.

However, the new formats and standards for retrieval systems for information storage and access require search adjustments, which may overload the results obtained and the relevance of the documents retrieved compared to the real needs of users.[7] This situation reinforces the value of representing the content of documents by cataloging subjects and their direct link with the search results, which can be improved by subject analysis. For that reason, an update is necessary.

Guidelines for subject analysis need to be formulated for indicative orientation of elements to be incorporated in the construction of standards. Due to studies on procedures, manuals identify these elements as essential for the analysis of subject cataloging in libraries.[8] On the other hand,

several investigations indicate that the adequacy and quality of the subject headings are essential for the processes of searching and retrieving information.[9]

This scientific research on subject analysis details methodological procedures, discusses and analyzes Individual Verbal Protocols and Ethnographic Research, and develops proposed guidelines for subject analysis.

Literature review: Knowing the documentary content

This research aims at studying the subject analysis in cataloging; however, due to the theoretical and methodological advances of indexing related to subject cataloging, this research considers the theoretical approach of both theories. Although they are different procedures, both understand that subject analysis is the first stage of analysis when reading, identifying, and selecting concepts in a document.

Thus, it is important to mention that indexing involves subject analysis to identify terms to represent what the document is about and its representation in any language of an indexing system, whereas subject cataloging involves subject analysis and is concerned with determining and describing the content of books and translating it into subject headings and classification notes.

Recent studies address research on subject cataloging and indexing as the differences between subject cataloging for books and indexing theses in university libraries. It was found that these processes are distinct since indexing presents a higher level of specificity.[10] Subject analysis of postcard images was investigated using the concepts of aboutness and ofness, and it was concluded that the concepts guided the search for the constituent elements of the images.[11] A study on the subject indexing literature in archives identified that the specificities and context of each document should be considered;[12] books with chapters by several authors were used as models of documentary reading for indexing and their applications in different contexts and document typologies.[13]

Authors have identified two stages in subject analysis in cataloging: conceptual analysis and translation.[14] Others also define five stages in subject analysis:[15] identification of the topics, identification of the author's intention, inference of the users, relation of the topics of subject with the collection, and relation of the topics of subject with documentary languages. This research, however, considers three stages:[16] documentary reading, identification of concepts, and selection of concepts.

Documentary reading focused on professional contexts is "the attitude of the reader toward the text, previously seen as a passive reception of messages; it considers the mental processing of information from comprehension and evolves to an interaction between the reader and the

text,"[17] recognizing that individuals "make use of all previous knowledge stored in memory"[18] for textual comprehension.

The concept identification stage occurs during reading comprehension, and documentary reading models for subject cataloging of scientific texts and books based on the exploration of the textual structure allows a cataloger to identify concepts by questioning.[19] When selecting concepts, the cataloger "describes the concepts identified in the document by subject cataloging terms, such as thesaurus, classification tables, etc."[20]

Table 1 discusses conceptual analysis methods:

Given the approaches described above, it is legitimate to understand that the variables must be considered because it is an activity carried out under various aspects, based on the cognition of the subject analyst with their previous knowledge and interpretative potential.

The variables of subject analysis involve knowledge of several levels related to professional cognition, technical aspects, and other types of knowledge that can improve understanding. This association can be made by semiotics, based on the synthetic nature of the elements originating from the representation of the subject.[21] Peirce's philosophical perception, and the interpretation in this type of intellectual activity carry along collateral experience and habits. The collateral experience comprises the experiences of the individual as well as the inferences of the cataloger in the representation of the subject and defines the habit as recurrent actions to "a predisposition to act."[22]

Table 1. Conceptual analysis methods.[23]

Approach	Identifying the author's intentions in writing the work (Purposive Method).
	Weighing the relative importance or prominence of elements in the work (Figure-Ground Method).
	Tallying references to various elements—an empirical approach (Objective Method).
	Selecting the themes or elements that hold the work together as a whole (i.e., what has been selected or included and what has been rejected or excluded) (Cohesion Method).
Approach	Examining the text:
	Scan the title, subtitle, author information, table of contents, and chapter headings;
	Read the introduction and book jacket;
	Sample the text, if necessary, or check book reviews;
	Determine if the work is homogenous or a composite work.
	Analyzing each unit:
	Determine the fundamental form of knowledge;
	Determine the discipline;
	Determine the topic;
	Determine the nature of the thought;
	Determine the nature of the text.
	Creating a Summary:
	Write a summary of the findings in natural language expressing the aboutness of the item.
Approach	Discipline;
	Topics;
	Names (of all types);
	Chronological elements;
	Form and/or Genre.

Other levels of knowledge are linked to understanding: textual linguistics, intertextuality, and documentary structure. In this sense, recognizing the composition and formation of each documentary typology that composes the collection means understanding and locating the most important information necessary for the representation of information with the guarantee that the documentary item is visible to the user.

Issues of textual linguistics contribute to the understanding of content:

> The cohesion and/or semantic coherence of the text [has] the objective of producing specific effects [of understanding]. Moreover, [...] coherence is not a property or quality of the text itself: a text is coherent for someone in a specific communication situation. To build coherence, someone must take into account not only the linguistic elements that make up the text, but also the encyclopedic knowledge, mutual knowledge and images, beliefs, convictions, attitudes, presuppositions, explicit or veiled intentions, immediate communicative situation, sociocultural context and so on.[24]

It is understood that cohesion and coherence for textual understanding are directly linked with other levels of knowledge, connecting with diverse areas and concepts.[25] Intertextuality can be presented in four topics[26]: thematic; stylistic; explicit and implicit; however, the last two provide textual comprehension related to other sources and ideas connected to the text, which reports us to previous knowledge as background for comprehension in subject analysis. In addition, the parameters for good development of subject analysis do not depend only on professional cognition or a normative methodology, it must have a procedures manual strategically elaborated as an institutional philosophy.[27]

The results of this discussion helped the elaboration of the categories of analysis of the Individual Verbal Protocol and Ethnographic Research: Collateral Experience and Habit;[28] Stages of the subject analysis;[29] Intertextuality;[30] and Textual Knowledge.[31]

Methodology

This research is characterized as theoretical-practical research, classified as exploratory research. Aiming at elaborating the guidelines for subject analysis, the methods used for observation were the Individual Verbal Protocol (IVP) and Ethnographic Research (ER) applied with the catalogers when performing subject analysis in subject cataloging.

The verbal protocol technique presents the individual and group modalities that are applied in different situations. The first consists of a task developed by a subject without the interference of the researcher; and the second modality comprises a social reading event to discuss a topic proposed for the meeting and can count on the participation of the researcher. Both modalities allow delimiting from initial application decisions to the form of analysis.[32]

For this research the Verbal Protocol was used to observe the cognitive processes of the cataloger in the development of the subject analysis based on "Think Aloud" that provides voice recording. The investigations continued with the EP, to raise reports, difficulties, observations, processes and actions of professionals, so that it was possible to make a parallel between the discussions obtained with the two techniques. It is important to clarify that ethnographic research was applied at the end of the IVP, as a retrospective interview that is already optionally planned after its application, to take up questions that were not clear.

Due to the complexity and large amount of data obtained from the IVP, the research universe comprised a sample of 16 university libraries, with a cataloger in each institution in the fields of Health Sciences, Exact and Earth Sciences, and Applied Social Sciences. As selection criteria, Brazilian and Spanish university libraries were chosen with more than 100,000 books in their collection. The sample was thus defined so that the analyses could provide the observation of the cataloging subjects in specific domains in university libraries of different cultures.

The instruments were applied in Brazilian and Spanish university libraries, with the aim at deepening the research on PVI and PE from the pre-established means described below (Table 2).

Individual Verbal Protocol (IVP)

The technique consists of "recording the lines that express the thoughts of the catalogers participating in the research while performing a certain task."[33] The application of IVP used the following procedures.[34]

Table 2. Definition of the survey sample.

Subjects	University Libraries (Capes Knowledge Areas)	Country	Quantity from University Libraries	Documentary item	Data collection instruments with each cataloguer
16 Subject analyst cataloguers	Medicine (Health Sciences)	Brazil	6	Indexing of a Book in each library	- Individual verbal protocol (PVI) e
	Computer Science (Exact and Earth Sciences)	Brazil	6		- Ethnographic Research (PE)
	Librarianship (Applied Social Sciences)	Spain	4		

Source: Prepared by the authors.

Procedures prior to data collection:

a. Research universe: definitions for data collection (see Table 2).
b. Basic text selection: books belonging to the specific domains of each area.
c. Task definition: the aim of the task was the cataloging subjects of books divided into the following stages of the subject analysis: documentary reading, identification, and selection of terms, which consisted of recording the cataloger's thoughts.
d. Cataloger selection: 16 catalogers from Brazilian and Spanish university libraries (see Table 2). The catalogers were selected according to their experience in cataloging, so we selected professionals with ten or more years of experience in cataloging in the indicated areas. Thus, the data obtained in the collection were essential to discuss the guidelines.
e. Informal conversation with all catalogers: present the research objectives and its importance for the development of knowledge organization.
f. Familiarization with the task of "think aloud": instructions to subjects.[35]

Procedures during data collection:

a. Recording of "think aloud" during the reading of the base text: the catalogers exteriorized their thoughts by reading the material out loud, generating the protocols of mental processes. The researcher controlled the digital recorder throughout the IVP application individually with each cataloger, as well as the informal conversation, familiarization with the IVP and instructions to the subjects.
b. Retrospective interview: Ethnographic research was applied with each of the catalogers soon after the IVP, with the aim of eliciting data obtained to clarify excerpts of "think aloud" considered inaccurate by the researcher.

Procedures after data collection:

a. Literal transcription of the catalogers' audio recordings: the transcribed text was defined according to the specific notations for transcription,[36] where each expression and feeling conveyed in the text is marked with a different symbol. The transcribed data from the Spanish catalogers were translated into Portuguese, so that they could be treated together with the transcribed data from the Brazilian catalogers. After the text was ready, it was translated into English.

Table 3. Thematic axes and categories of analysis.

Thematic axes	Categories of analyses	Sources
1 Subject analysis	Stages of subject analysis	Authors[37]
	Quality and Evaluation of Subject Analysis	Authors[38] / Collections - IVP and PE
2 Central conceptual areas	Information unit planning and goals	IVP and PE collections
	Delimitation of the domain area	IVP and PE collections
	Conceptual aspects related to the subject analysis	IVP and PE collections
3 Cognitive Factors	Representation instruments	IVP and PE collections
	Representation levels	Author[39]/ IVP and PE collections
	Documentary structure	IVP and PE collections
	Textual knowledge	Author[40]
	Intertextuality	Author[41] / IVP and PE collections
	Habits and collateral experience	Author[42]
	Proficiency	IVP and PE collections
4 Users	Public user	IVP and PE collections
	Access to user-controlled language	IVP and PE collections

Source: Prepared by the authors.

b. Detailed reading of data to search significant and recurrent phenomena to construct the categories of analysis: detailed reading of the transcriptions.

c. Building categories of analysis: developed following the transcriptions of IVPs and ERs, and the theory studied (see Table 3).

d. Returning to the data to take excerpts from the discussion that exemplify each category of analysis: analysis of protocols for observation of the cataloger's reports.

Ethnographic research (ER)

Ethnographic research "aims to discover new relationships and new ways of understanding reality, which is why it seeks the formulation of hypotheses, concepts and explanations and not their testing."[43] In this investigation, ER was used as the elicitor of the data collected with the individual verbal protocol to obtain the retrospection of the catalogers on the cataloging subjects performed.

Ethnographic research is composed of three phases:[44]

a. Participant observation: the observation of the subject occurred with the analysis of the protocols recorded and transcribed; and the discussions were based on the cataloger's knowledge on the context to which they belong, on the users, the language, procedures manual, any contextual aspect that directly influences the analysis of the subject, and questions that involved the retrospective interview.

b. Interviews: the interviews were tailored to the reality of each cataloger, according to the responses obtained in the IVP the researcher was

allowed to determine the problems and difficulties, mainly related to the professional's daily life, to contrast with the previous ones, cataloger's verbalizations (IVP) with their new elucidation (ER).

c. Analysis: They were based on the transcriptions of the interviews based on the same logic of analysis described in the procedures after the IVP data collection: building categories of analysis and returning to data to remove sections of the discussion that exemplify each phenomenon.

Results of the analysis of Individual Verbal Protocols and Introspective Research

The responses from the catalogers of university libraries in Brazil and Spain on the two techniques of data collection were analyzed in categories elaborated according to the literature of the area and the transcriptions of IVP and ER and grouped in four thematic axes to obtain better discussion of data. The principle applied in the development of these four thematic axes followed the common aspects found in the categories of analysis.

Table 3 illustrates the distribution of the categories of analyses among the four thematic axes and the sources of preparation of these categories.

This structure of analysis made it possible to discuss the transcribed data from Individual Verbal Protocols about the professional practice, and to identify the strategies and skills that portrayed the cataloger's cognition about book cataloging, and thus obtained representations of the catalogers' mental processes. In the other hand, the analysis with the ER transcriptions revealed some contradictions and explanations of the responses achieved with the application of IVPs.

Both methods proved to be pertinent for a more realistic observation of the subject analysis in cataloging, providing a comparative and complementary discussion between the occurrences obtained by the practice of subject analysis (IVP) and the explanation of the catalogers about the unfolding of their own experience with subject analysis (ER).

The 14 categories of analysis were composed of excerpts from IVP and ER transcriptions and critically discussed, what made it possible to cross results between the observation of the practice of subject analysis and the subject analyst report.

The distribution of the fourteen categories in the four thematic axes generated the guidelines for the analysis of the subject. Thus, all analyzes followed the same discussion structure as the example of the documentary analysis category. It is important to mention that in this section some of the main catalogers' speeches were highlighted in some categories:

Stages of subject analysis:

Documentary reading:

For instance, regarding the documentary reading category, IVP revealed that two catalogers from the Brazilian libraries consider reading the information on the cover and the summary as essential parts to find out the subject of the book. However, when questioned about the relevance of documentary reading for representation, both agreed that this procedure is carried out with attention to excerpts of the document, where they will have more understanding of the subject. Two Spanish catalogers highlighted the importance of documentary reading; however, during the retrospective interview, one of them stated that lack of time is a barrier to the completion of this process; and according to the second cataloger, it can be done only by the title of the book, with some exceptions using the prologue and the content of the book.

#Transcriptions of the lines of the catalogers:

a. The comprehension of the reading technique is based on the professional's prior knowledge.
b. The previous knowledge of the cataloger influences the decision-making.
c. Technical reading is an important prerequisite for cataloging subjects.
d. In books with chapters by different authors, it is common that the chapters contain different themes.
e. It is not recommended to identify the main subject of the book only by its title.
f. The titles of books on Exact and Earth Sciences and Health Sciences, in most cases, do not allow for various interpretations. It is more common for books to deal with a single subject.
g. The books on Applied Social Sciences may give room for interpretation. For example: books on sociology may be in Economic Sciences, Social Sciences, or Education.
h. Recognize the structure of the book and where the important information is.

Identification of concepts:

#Transcriptions of the catalogers' lines:

a. Identifying terms from the documentary content is an important step, as it will help the cataloger to validate the cataloging subject terms retrieved in cooperative cataloging.
b. Consult experts.
c. Insert representative concepts that match the text, context, and user.
d. Identify the main idea of each chapter.
e. Negative habits influence the decision-making.

f. Books of Applied Social Sciences in general are more subjective, so the identification of concepts may take longer.
g. Health Sciences terminological updating is dynamic.
h. The Health Sciences have their own acronyms to represent the terms.
i. In the Exact and Earth Sciences, the series are well delimited and are usually thematic.
j. The identification of the topic is faster in the Exact and Earth Sciences and Health Sciences, as they are more direct and objective.
k. Avoid meaningless/empty words.

Selection of concepts:
#*Transcriptions of the catalogers' speeches:*

a. The selection will start from the previous identification of the terms.
b. Know in detail the documentary language used in the library.
c. It is important that the selected terms are in accordance with the standards, controlled language, users, and context of the information unit.
d. Know the interests of the library users.
e. The terms will necessarily be validated with a documentary language.
f. Verify that the identified terms are authorized.
g. Periodic meetings with the other catalogers, classifiers, and indexers are important so that specific doubts are taken to the group.
h. In situations where the controlled language is not specific, suggestions about new subject authorities should be sent so that it can be evaluated in meetings for quality control (according to the standards of each library).

Quality and Evaluation of Subject Analysis:
#*Transcriptions of the catalogers' speeches:*

a. The quality of the subject analysis will be observed from the pertinence of the recovered documents.
b. The evaluation is necessary since the terms described in the representation must correspond to the user's search strategies.

Representation levels

Exhaustiveness:
#*Transcriptions of the cataloger's lines:*

a. Exhaustiveness is an essential element to be addressed in the procedures manual, where this manual should address updating to subject headings.

Specificity:
#Transcriptions of the cataloger's lines:

a. Specificity is an essential element to be addressed in the procedures manual.
b. The updating of the documentary language will allow better levels of specificity.

Relevance:
#Transcriptions of the catalogers' speeches:

a. Feedback in the recovery (user study).

Documentary structure:
#Transcriptions of the catalogers' lines:

Example 1 – Books: Books do not maintain an internal linear structure, however, the cataloger's experience in using this type of documentary item will help him to distinguish the most common aspects present in books from areas of knowledge.

Example 2 – Scientific articles: Unlike books, articles have a common structural division between different areas of knowledge, mainly because most journals stipulate similar standards of presentation: title, abstract, keywords, introduction, applied theory, methodology, results, final considerations, and references.

Example 3 – Theses, dissertations, and Conclusion Papers: textual structure similar to that described in scientific journal articles, with the addition of pre- and post-textual elements.

Intertextuality:
#Transcriptions of the catalogers' lines:

a. In books of Applied Social Sciences the texts can be intertwined in a network of multidisciplinary knowledge, requiring the subject analyst to have extensive interdisciplinary knowledge to obtain understanding.

Habits and collateral experience
#Transcriptions of the cataloger's speeches:

a. Continued education can directly influence the positive incidences in the cataloger's decision-making.
b. Establish strategies from the literature and methodologies of the area.

Proficiency:
#Transcriptions of the catalogers' speeches:

a. The training courses are effective practices for the improvement and theoretical and methodological deepening in the library areas of domain.
b. Consultation with specialists.

Public user:
#Transcriptions of the catalogers' speeches:

a. It is necessary to have information on actual and potential users, research groups, and specialists of the information unit so that the representation can fulfill its social role for access.

Access to user-controlled language:
#Transcriptions of the catalogers' lines:

a. The access to the language by the user can be a way to solve the lack of updating of controlled languages because you can check the reference descriptors in the areas of interest.

Subject analysis guidelines for book subject cataloging

The subject analysis guidelines listed below were formulated with the results of the four thematic axes and the categories of analysis that comprise them. Table 4 shows the correspondence between the thematic axes and the sections of the guidelines as well as the subsections originating from the analysis categories:

The four sections are composed of subsections, a description, and the purpose for which it was developed. The subsections describe the procedural steps for the positive development of the subject analysis, as well as excerpts with explanatory purpose transcribed from the speeches of the catalogers about the important aspects corresponding to their predominant performance in IVP and ER. The discussions presented in this item were based on the areas of the libraries studied in this research.

Section I – Main guidelines that libraries need to consider in the elaboration of procedures manual for subject cataloging.

Description and objective of the section: The first section refers to the macro knowledge that the cataloger needs to acquire about the analysis of the documentary subject, based on concrete parameters about the unit of information, the collection, and the theoretical and methodological knowledge in the area.

Table 4. Formulations of the sections of the subject analysis guidelines and the thematic axes.

Thematic axes	Sections of the subject analysis guidelines	Subsections of the subject analysis guidelines
1 Central conceptual areas	Section I – Main guidelines that information units need to consider when developing an Indexing Policy	Information unit planning and goals Delimitation of the domain area Conceptual aspects related to the subject analysis
2 Cognitive factors	Section II – Factors related to the cataloger cognition for understanding the variables of the subject analysis	Instruments of representation Levels of representation Exhaustiveness Specificity Relevance Recognize the structure of documents for document reading and textual understanding Textual knowledge Cohesion and textual coherence Intertextuality Habits Collateral experience Proficiency
3 Users	Section III – Information about the users of the information unit	Public user Access to user-controlled language
4 Subject analysis	Section IV – Specific Procedures for Subject Analysis	Stages of subject analysis Documentary reading Identification of concepts Concept selection Quality and Evaluation of Subject Analysis Quality Evaluation

Source: Prepared by the authors.

It is composed of the planning and goals of the unit of information; delimitation of the area of domain; and conceptual aspects related to the analysis of the subject.

Information Unit Planning and Goals: Outlining and understanding the objectives and goals of the organization will help in the development of the procedure's manual.

Delimitation of the area of domain: For good articulation and thematic understanding it is suggested that the subject analyst domain the areas he will be in charge of representing the information. A good measure of this understanding is the feedback that users will give to the system. For this purpose, it is essential to be in contact with the Reference Sector.

Conceptual aspects related to the subject analysis: The understanding of the theoretical and methodological contributions on subject cataloging, subject analysis, alphabetic documentary languages, the procedures manual, and information retrieval will bring quality to the representation by subject.

Section II – Factors related to the cataloger's cognition for understanding the variables of the subject analysis

Description and purpose of the section: The decisions for subject analysis parameters are not neutral and depend on the cataloger's cognition, therefore, this section includes variables that can influence this decision process. It is composed of: instruments of representation; levels of representation; recognition of the structure of documents for documentary reading and textual understanding; textual knowledge; habits; and collateral experience and proficiency.

Instruments of representation: It is important to master the instruments and resources available for representation, mainly for the elaboration of the institution procedures manual; the controlled language and its specifications, and the functioning of the cooperation system between libraries before verifying data retrieved.

Levels of representation: Specificity and completeness are elements that are related to the relevance of the representation of information in relation to the user's search needs. Exhaustiveness: "Exhaustiveness [determines a sufficient number] of terms to cover the thematic content of the document in a fairly complete manner."[45]

Specificity: "It implies the use of a much smaller number of terms in order to cover only the main thematic content of the document."[46]

Relevance: The user's study makes it possible to investigate the adequacy between the concepts described in the representation and the informational needs of the user.

Recognize the structure of documents for documentary reading and textual understanding: the greater the familiarity with documentary typologies, the greater the mastery in reading.

Textual knowledge: The recognition of the structures that make up the text allows us to understand and identify content concepts that represent it.

Textual cohesion and coherence: The understanding of cohesion and textual coherence will make it possible to understand the communicative purpose of the text.[47] Intertextuality: Intertextuality provides connections with other areas and concepts.[48] The understanding of explicit and implicit intertextuality[49] gives meaning to intertext/text.

Habits: Negative habits in subject analysis: difficulty in reading documents for understanding; use of cooperative cataloging records without data validation as a definitive step for the identification and selection of concepts.

However, the break of habits allows new reflections about reading for the identification and selection of concepts in the subject analysis and use of methodological instruments. Furthermore, it allows adapting and validating the metadata obtained in the records of the cooperative cataloging according to the need for real demand.

Collateral experience: The habit will guide the cataloger's decision, and the results will always be present in the representation of the information.

Proficiency: There are two premises about obtaining proficiency: the first is that the subject analyst will develop his own techniques to improve the representation from the domain of the field; and the second may cause misinformation on the subject analysis.

Section III – Information about the users of the information unit

Description and objective of the section: The needs of library users are indicators for both the decision-making process of the subject analysis, and for feedback in the formulation of search strategies. The section is composed of the user's audience and the access to the language controlled by the user.

User's audience: knowing and defining the profile and interests of the user audience according to the objectives and demands of each institution enable targeted and more precise results in the representation of information.

Access to user-controlled language: Providing users with knowledge of controlled language is an important strategy to minimize the problem of searches by subject in catalogs, especially in remote access.

Section IV – Specific Procedures for Subject Analysis

Description and objective of the section: This section gathers important aspects related to the subject analysis, drawing attention to the quality and evaluation of this process. It is composed by the steps, quality, and evaluation of the subject analysis.

Stages of the subject analysis: The analysis of catalogers' procedures, strategies, and difficulties in the subject analysis allowed a compilation of important aspects and directions for the documentary reading, identification, and selection of concepts.

Documentary reading: It is indicated to cautiously analyze the excerpts containing the most significant information of the document.

Identification of concepts: The identification of concepts occurs when reading texts and models of documentary reading.[50]

Concept selection: The selection of concepts will occur according to the procedure's manual developed in the information unit and integrated to the two previous stages.

Quality and Evaluation of Subject Analysis. Quality: The quality of the analysis is subject to several factors: indexer (professional cognition); vocabulary (standards of textual construction and specific instruments); environmental (external conditions); document (structure and factors that constitute the document); process (norms and manuals for cataloging subjects);[51] habits; collateral experience; users (relevance of representation).

Evaluation: The evaluation of the representation will verify its applicability within the information retrieval system. The evaluation can occur through qualitative intrinsic evaluation; quantitative intrinsic evaluation; extrinsic evaluation through interconsistency; and extrinsic evaluation through retrieval.[52]

Final considerations

The dynamics of access to catalogs require the incorporation of metadata representation in a consistent and effective way so that the search by subject is contemplated positively in systems of information retrieval both in physical and digital format.

The research presented a detailed look at several variables that make up subject analysis in subject cataloging. The proposal presented aimed at a methodological perspective to discuss the improvement of quality in the development of subject analysis in subject cataloging through the identification of concepts to be applied in libraries.

Above all, the use and confirmation of the verbal protocol as a cognitive method for the recognition of the strategies and difficulties encountered by catalogers made possible the in-depth analysis of the entire procedure. It is important to highlight that the verbal protocol in Information Science has been used only for studies aimed at cataloging subjects revealing its effectiveness for reading with the objective of representing content.

In this scenario, discussing and presenting the guidelines for subject analysis in the subject cataloging proved to be a challenge for defining general guidelines for a broader purpose of improvement and indicators that compose the subject analysis, control of its variables according to a professional cognitive perception and relevant aspects that the libraries need to consider for the construction, and incorporation as an essential component in the norms and in the subject cataloging procedures manual.

Author contributions

Franciele Marques Redigolo: Conceptualization, Methodology, Investigation, Writing – Review & Editing, Project administration; Mariângela Spotti Lopes Fujita: Conceptualization, Methodology, Writing – Review & Editing, Project administration; Isidoro Gil-Leiva: Conceptualization, Methodology, Writing – Review & Editing.

Funding

ORCID

Franciele Marques Redigolo ⒾⒹ http://orcid.org/0000-0001-6277-2960
Mariângela Spotti Lopes Fujita ⒾⒹ http://orcid.org/0000-0002-8239-7114
Isidoro Gil-Leiva ⒾⒹ http://orcid.org/0000-0002-7175-3099

Notes

1. Daniela Majorie Akama dos Reis, "A leitura documentária de bibliotecários jurídicos: um estudo realizado a partir de aspectos da semiose e teoria da inferência observados na estrutura textual de doutrina" (PhD thesis, Universidade Estadual Paulista, 2019), https://repositorio.unesp.br/bitstream/handle/11449/181849/reis_dma_dr_mar.pdf?sequence=3&isAllowed=y
2. Arlene G. Taylor and Daniel N. Joudrey, "On Teaching Subject Cataloging," *Cataloging & Classification Quarterly* 34, no. 1-2 (2002): 221–30. https://www.tandfonline.com/doi/abs/10.1300/J104v34n01_13
3. Ibid., 2.
4. Ibid., 2.
5. Ibid., 2.
6. Ibid., 2.
7. Hope A. Olson and John J. Boll, *Subject Analysis in Online Catalogs*, 2nd ed. (Colorado: Libraries Unlimited, 2001), 333.
8. Mariângela S. Lopes Fujita, "Políticas de organização e representação da informação" (CAPES: UAB, Rio de Janeiro [Departamento de Biblioteconomia, FACC/UFRJ], 2019).
9. Jung-Ran Park and Yuji Tosaka, "Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms," *Cataloging & Classification Quarterly* 48, no. 8 (2010): 696–715. https://www.tandfonline.com/doi/abs/10.1080/01639374.2010.508711; Ravit H. David and Dana Thomas, "Assessing Metadata and Controlling Quality in Scholarly Ebooks," *Cataloging & Classification Quarterly* 53, no. 7 (2015): 801–24. https://www.tandfonline.com/doi/abs/10.1080/01639374.2015.1018397; Weijing Yuan, Marlene Van Ballegooie, and Jennifer L. Robertson, "Ebooks Versus Print Books: Format Preferences in an Academic Library," *Collection Management* 43, no. 1 (2018): 28–48. https://www.tandfonline.com/doi/abs/10.1080/01462679.2017.1365264; Isidoro Gil Leiva et al., "The Abandonment of the Assignment of Subject Headings and Classification Codes in University Libraries due to the Massive Emergence of Electronic Books," *Knowledge Organization* 47, no. 8, (2020): 646–67. http://eprints.rclis.org/40981/1/Gil-LeivaFujitaDiazMajorie2020.pdf
10. Roberta C. D. E. Tartarotti and Mariângela S. Lopes Fujita, "A catalogação de assunto e a indexação em bibliotecas universitárias: um estudo comparativo com protocolo verbal," *Scire* 23, no. 1 (2017): 57–66. http://hdl.handle.net/11449/232621
11. Nina Cláudia Mendonça Campos de Miranda, "Análise de assunto das imagens dos cartões-postais do acervo Alaíde Lisboa de Oliveira: o uso do aboutness e do ofness na extração dos conceitos" (PhD diss., Escola de Ciência da Informação da Universidade Federal de Minas Gerais, 2019).
12. Graziela Martins de Medeiros and Marisa Bräscher Basílio Medeiros, "A indexação de assunto em documentos arquivísticos," *Revista Digital de Biblioteconomia & Ciência da Informação* 18 (2020): 1–28. https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8657824/22148
13. Mariângela S. Lopes Fujita, Roberta V. Alves, and Carlos C. Almeida, orgs., *Modelos de leitura documentária para indexação*: *abordagens teóricas interdisciplinares e aplicações em diferentes tipos de documentos* (Marília: Oficina Universitária; Cultura Acadêmica, 2020).
14. F. W. Lancaster, *Indexação e resumos*: *teoria e prática*, 2nd ed. (Brasília, DF: Briquet de Lemos, 2004); A. G. Taylor, *The Organization of Information* (Englewood, CO: Libraries Unlimited, 1999).

15. Alenka Sauperl and Jerry D. Saye, "Subject Determination during the Cataloging Process: An Intensive Study of Five Catalogers," in *Proceedings of the 9th ASIS SIG/CR Classification Research Workshop* (Silver Spring, MD: American Society for Information Science, 1998), 119–38. https://journals.lib.washington.edu/index.php/acro/article/view/12748

16. Associação Brasileira de Normas Técnicas 12676, *Métodos para análise de documentos—determinação de seus assuntos e seleção de termos de indexação* (Rio de Janeiro: Abnt Nbr, 1992); Eduardo Wence Dias and Madalena Martins Lopes Naves, *Análise de assunto*: *teoria e prática*, 2nd ed. (Brasília: Thesaurus, 2013); Mariângela S. Lopes Fujita, *A análise documentária no tratamento da informação*: *as operações e os aspectos conceituais interdisciplinares* (Marília, São Paulo, Brasil: Departamento de Ciência da Informação, FFC/UNESP, CNPq, 2003); Mariângela S. Lopes Fujita, *A leitura documentária na formação inicial do indexador: a abordagem sociocognitiva na investigação de estratégias de ensino* (Marília, São Paulo, Brasil: Departamento de Ciência da Informação, FFC/UNESP, CNPq, 2004).

17. Ibid., 16.

18. Dulce Amélia de Brito Neves, Eduardo Wense Dias, and Ângela Maria Vieira Pinheiro, "Uso de estratégias metacognitivas na leitura do indexador," *Ciência da Informação* 35, no. 3, (2006): 141–52. https://www.scielo.br/pdf/ci/v35n3/v35n3a14.pdf

19. Mariângela S. Lopes Fujita, *A análise documentária no tratamento da informação*: *as operações e os aspectos conceituais interdisciplinares* (Marília, São Paulo, Brasil: Departamento de Ciência da Informação, FFC/UNESP, CNPq, 2003).

20. Ibid., 16.

21. Carlos C. de Almeida, "Sobre o pensamento de Peirce e a Organização da Informação," in *Proceedings of the XI Encontro Nacional de Pesquisa em Ciência da Informação, Rio de Janeiro, RJ, Brasil* (2010).

22. Franciele Marques Redigolo and Carlos Cândido de Almeida, "Algumas contribuições da perspectiva filosófico semiótica de Peirce para a análise de assunto," *DataGrama-Zero* 13, no. 3 (2012): 1–12. https://brapci.inf.br/index.php/res/v/7848.

23. R.M. Holley and D.N. Joudrey, D. N. Aboutness and Conceptual Analysis: A Review. *Cataloging & Classification Quarterly* (2020). https://www.tandfonline.com/doi/abs/10.1080/01639374.2020.1856992?journalCode=wccq20.

24. Ingedore G.Villaça Koch, *O texto e a construção dos sentidos*, 10th ed. (São Paulo: Contexto, 2014).

25. Ingedore G.Villaça Koch, *Desvendando os segredos do texto*, 7th ed. (São Paulo: Cortez, 2013).

26. Anna Christina Bentes, Ingedore G. Villaça Koch, and Mônica Magalhães Cavalcante, *Intertextualidade: diálogos possíveis* (São Paulo: Cortez, 2007).

27. Ibid., 8.

28. Ibid., 23.

29. Ibid., 18, 19.

30. Ibid., 25, 26.

31. Ibid., 24.

32. Adolfo Alonso-Arroyo et al., "Protocolo verbal: análisis de la producción científica, 1941-2013." *Informação & Sociedade: Estudos* 26, no. 2 (2016): 61–76. https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/28182

33. Tais Caroline L. Mattos et al., "Protocolo verbal: verbalizações concorrente e retrospectiva," *Revista Ibero-Americana de Ciência da Informação* 10 (2017): 49–66. https://periodicos.unb.br/index.php/RICI/article/view/2475

34. Mariângela S. Lopes Fujita, org., et al. *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais [online]* (Marília: Editora UNESP – Cultura Acadêmica, 2009).

35. M. I. A. Nardi, "As expressões metafóricas na compreensão de texto escrito em língua estrangeira" (PhD thesis, Pontifícia Universidade Católica, 1993).

36. Mônica Magalhães Cavalcante, *I-n-t-e-r-a-ç-ã-o leitor-texto: aspectos de interpretação pragmática* (Campinas: UNICAMP, 1989).

37. Associação Brasileira de Normas Técnicas 12676, Métodos para análise de documentos-determinação de seus assuntos e seleção de termos de indexação (Rio de Janeiro: Abnt Nbr, 1992).

38. F.W. Lancaster. *Indexação e resumos*: *teoria e prática*. (2nd ed.). Briquet de Lemos. 2004. A.G. Taylor. *The organization of Information*. Libraries Unlimited. 1999. Isidoro Gil-Leiva, I. *Manual de indización*: *teoría y práctica*. Trea. 2008

39. F.W. Lancaster. *Indexação e resumos*: *teoria e prática*. (2nd ed.). Briquet de Lemos. 2004.

40. Ingedore G. Villaça Koch (2014). *O texto e a construção dos sentidos*. (10th ed.). Contexto.

41. Anna Christina Bentes, Ingedore G. Villaça Koch, and Mônica Magalhães Cavalcante, Intertextualidade: diálogos possíveis (São Paulo: Cortez, 2007).

42. Franciele Marques Redigolo and Carlos Cândido de Almeida, "Algumas contribuições da perspectiva filosófico semiótica de Peirce para a análise de assunto," *Data-Grama-Zero* 13, no. 3 (2012): 1–12. https://brapci.inf.br/index.php/res/v/7848.

43. Graziela Zambão Abdian Maia, "Pesquisa etnográfica e estudo de caso," in *Pesquisa em educação: passo a passo*, ed. A. C. F. B. Labegalini (Marília: Edições M3T, 2007).

44. Marli Elisa D. A. André, *Etnografia da prática escolar* (Campinas: Papirus, 1995).

45. Ibid., 14.

46. Ibid., 14.

47. Ingedore G.Villaça Koch, *A coesão textual* (São Paulo: Contexto, 2010).

48. Ibid., 30.

49. Ibid., 25.

50. Ibid., 16, 19; ISO 5963:1985, *Documentation: Methods for Examining Documents, Determining their Subjects, and Selecting Indexing Terms* (Geneva: International Organization for Standardization, 1985); José Antonio Moreiro González, *El contenido de los documentos textuales: su análisis y representación mediante el lenguaje natural* (Asturias: Trea, 2004).

51. Ibid., 14.

52. Ibid Isidoro Gil-Leiva, I. Manual de indización: teoría y práctica (Trea, 2008).