Brian Quinn
Floral Park, New York

Recent Theoretical Approaches in Classification and Indexing

Brian Quinn is a recent graduate from the Graduate School of Library and Information Science at the University of Illinois. His interests are in classification theory, subject analysis, and social sciences librarianship. He is currently engaged in research on teamwork in academic libraries.

Quinn, B.: **Recent theoretical approaches in classification and indexing.**
Knowl.Org. 21(1994)No.3, p.140-147, 40 refs.
This article is a selective review of recent studies in classification and indexing theory. A number of important problems are discussed, including subjectivity versus objectivity, theories of indexing, the theoretical role of automation, and theoretical approaches to a universal classification scheme. Interestingly, much of the work appears to have been done outside the United States. After reviewing the theoretical work itself, some possible reasons for the non-American origins of the work are explored.
(Author)

1. Introduction

This study is a selective review of some of the more interesting theoretical approaches in classification and indexing. As such, it is not meant to be a comprehensive or exhaustive survey of every development currently taking place. Its purpose is simply to call attention to some of most interesting work that has recently been done in the field.

The majority of this work does not appear to have been done in the United States. Of course, there are a number of theoreticians writing in the U.S., such as Dagobert Soergel, Elaine Svenonius, Jean Perreault, and Francis Miksa. And historically the U.S. has produced Dewey, Cutter, Bliss, and more recently individuals like, Lubetzky, Dunkin, Immroth, Richmond, and Painter.

Nonetheless, the relative lack of American contributions is enough to give one cause for reflection. What is it about American classification in particular, and the American mind in general, that makes it so pragmatic and practice-oriented? After outlining some recent theoretical approaches in classification, this study concludes with speculating about some of the possible reasons why heory seems to be less central in American classification.

2. Subjectivity versus Objectivity

Among the more interesting theoretical issues that catalogers are trying to resolve concerns the degree of subjectivity vs. objectivity in classification. Researchers are trying to determine if the inquirer's and indexer's point of view plays a role in the ways he classifies information for retrieval (1, p.204). Some of the more extreme subjectivists verge on mentalism in holding that indexing is an unsolvable problem in information science. They believe that definitions of concepts are in the eyes of the beholder rather than in what is beheld.

The work of the individual indexer, for example, is highly subjective and open to personal interpretation because the decisions he makes involve judgements of the value of what is presented. His interpretations may or may not be the same ones that occur to the inquirer in the process of searching. Thus a subjective bias is introduced in the classification process.

Subjectivity is also a consideration with regard to the searcher. How to incorporate knowledge about users in designing an effective information retrieval system is viewed as being of major importance. So far, information retrieval systems have been structured to deliver the same response regardless of the user's cognitive characteristics. Those working in this area suggest that a more flexible system capable of adapting to different users is needed. The object is to create a retrieval system fluid enough to match the fluidity of the searching process of the unique and fluid inquirer (2. p.63).

An opposing group of researchers are critical of this subjective approach and argue for a more objective view. They have applied this view to the practice of indexing. Rather than attempting to discover the unconscious mental processes by which people derive indexing phrases from texts, efforts should focus on examining the objective rules they follow (3, p.94). Rules are not subjective or mysterious, they reason, but are actually social conventions in the form of practices, customs, or techniques. It is rules, not hidden mental processes, that people use to derive indexing phrases from texts.

The goal of indexing theory thus becomes not one of discovering subliminal processes but of constructing explicit, well formulated rules that can be used to yield indexing phrases from texts. The problems of indexer inconsistency is not solved by discovering cognitive functioning or by bringing order to the variety of tacitly known rules unconsciously followed by indexers. The solution lies in replacing vague rules subject to indexer interpretation, with more precise rules that establish clear standards of correctness. Experts need to reach a consensus as to what rules indexers should follow and how their performance is to be evaluated. It is the structural properties of the text itself, not the mentnl rules of text processing, that yield criteria of significance for the construction of indexing phrases.

This is an additional advantage to regarding rules as social conventions or constructs, the objectivists argue. One can really only understand the rule by placing it in its social context, including the historical, economic, political, and

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

cultural domains. Rules are not ahistorical, classless, or genderless, and are therefore subject to critical inquiry and vulnerable to social bias. This approach brings into prominence the social role of retrieval practices. The relevance of the subjective/objective debate in classification theory can be seen in de Grolier's work (4, p.64). He emerges on the objectivist side in proposing the study of the semantic structure of natural languages as a means of devising an intermediary language to interconnect various natural and information retrieval languages. This necessitates an objective basis for the organization of knowledge. De Grolier suggests that Bliss's 1929 theory of scientific and educational consensus provides the foundation for further attempts to furnish an objective basis for the systemization of knowledge. By emphasizing the social character of all classifications, he paved the way for an emerging relativism.

More recently, de Grolier cites studies of the linkages between the fields of science and technology as revealed by the Science Citation Index. The listing of journals under more than one subject category indicates strong linkages between agriculture and botany, or psychology, neurology, and psychiatry. These linkages, de Grolier suggests, demonstrate that "intuitively reasonable" sequences have actual objective nature.

3. Toward a Theory of Indexing

Another area of classification research that is the subject of considerable interest abroad is indexing theory. The lack of an indexing theory to explain the indexing process is a major blind spot in classification. According to one researcher, an indexing theory should consist of five levels. The first is concordance, which consists of references to all words in the original text arranged in alphabetical order. The second is the information theoretic level, which calculates the likelihood of a word being chosen for indexing based on its frequency of occurrence within a text. For example, the more frequently a word appears, the less likely it is to be selected because the indexer reasons the document is "all about that."

The third level is the linguistic one. This level of indexing theory attempts to explain how meaningful words are extracted from large units of text. Indexers regard some of a document as especially rewarding. Opening paragraphs, chapters, or sections, and opening and closing sentences of paragraphs are more likely to be a source of indexable words, as are definitions.

Beyond individual words or phrases lies the fourth level, the textual or skeletal framework. When an author creates a work, he does so in an organized manner which produces a skeletal structure clothed in text. The successful indexer needs to disinter this skeleton by searching for clues on the surface. Certain markers can be identified which hold a text together and accentuate its key elements.

The fifth and final level of indexing theory is the inferential level. An indexer is able to make inferences about the relationships between words or phrases by observing the paragraph and sentence structure, and stripping the sentence of extraneous detail. This inferential level makes it possible for the indexer to operate in novel subject areas (5, p.121).

The number five also plays a pivotal role in the indexing theory proposed by Robert Fugmann. His theory is based on five general axioms, which he claims have obvious validity and are in need of no proof. He believes they explain all currently known phenomena in information supply.

The first axiom is the *axiom of definability*. Compiling information relevant to a topic can only be accomplished to the degree to which a topic can be defined. The second axiom is the *axiom of order*, with Fugmann explaining that any compilation of information relevant to a topic is an order creating process.

The third axiom, called the *axiom of the sufficient degree of order*, posits that the demands made on the degree of order increase as the size of a collection and frequency of the searches increase. This is followed by the *axiom of predictability*, which says that the success of any directed search for relevant information hinges on how readily predictable or reconstructible are the modes of expression for concepts and statements in the search file. The fifth *axiom of fidelity* equates the success of any directed search for relevant information with the fidelity with which concepts and statements are expressed in the search file (6, p.13).

Fugmann has also taken up the difficult question of *indexing consistency*. Many investigators working in this area believe that indexing consistency should be somehow related to indexing quality and search effectiveness. Yet others have pointed out that consistent indexing can be consistently bad.

Even though perfect consistency can be readily achieved by the literal extraction of natural language expressions used by the author of the work, it is of little use to the searcher. Neither the searcher nor the retrieval system can foresee which words, phrases, or expressions the author has used and the indexer has chosen, and thus which search terms to use. Unaltered extraction of text words may actually be less conducive to effective retrieval than a more mediated process whereby the indexer attempts to select the words he suspects the user might use Consistency in the form of a single mode of expression can have a limiting effect on user access. This makes it much less essential to indexing quality than predictability. The user must be able to predict the terms chosen for indexing. Consistency is only essential in the initial selection of key concepts.

Fugmann believes that the real purpose of controlled vocabulary and classification in general is to enhance representational predictability. This is why he made it his fourth axiom. The element of predictability is largely missing from natural language expressions, particularly the more general ones. Terminological consistency is likewise next to nil due to the infinite number of paraphrasal forms in which they may appear in the text. Predictability is a much more important factor in contributing to overall indexing quality than consistency (7, p.21).

Another common belief among researchers that Fugmann takes issue with is the value of automation in indexing. Much of contemporary information science research has focused on improving methods for algorithmic processing of natural language texts. So much so that some individuals are now

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

claiming that this process is superior to human indexing. The author points out that these claims disregard the fact that human indexing can also be improved and that its potential is still far from being realized. One advantage of human cognition is its ability to view the same word or phrase quite differently depending upon the context. "Benzene" means one thing to a chemist, but quite another to a fireman.

Recognizing these different meanings constitutes a virtually insurmountable obstacle to developing purely algorithmic methods to a near human level of perfection. Only humans are capable of recognizing the equivalences of paraphrases and of lexical expressions such as descriptors. Fugmann is convinced the greatest potential for automated systems is in serving as aids to human indexers (8, p.65).

4. The Theoretical Role of Automation in Classification

It should be noted that this rather ambivalent view of computers is hardly typical of classification theorists. Many are quite enamored with the potential of automation to offer a radically new and different approach to classification. Computers allow information to be arranged and accessed in ways that will underscore the interconnectedness of different fields and systems of knowledge.

For example, prior to the development of computers, a library that arranged its holdings according to Dewey was not able to use LC classification. But a computer is multilinear and hence not limited to any particular set of relations. It allows books to remain in Dewey order while simultaneously making them accessible via different classification systems. This allows the possibility of comparing how the altemate systems of Bliss or Ranganathan use different concepts to classify a given item.

The computer is capable of changing classification in other ways. In large libraries, computers are able to catalog all personal names and their variations together to be accessed from a single source. This is much easier than having to search a multi-volume catalog of names to see which variant a particular name might be listed under.

The full text storage capabilities of online systems further reduce the possibility that books on a specific subject will be physically classed apart. Computers will help to synthesize previously fragmented disciplines by establishing new connections between different fields and subfields. By providing us with a bigger picture, they are helping to extend the limits of knowledge (9, p.11).

One way computers are able to help synthesize fragmented disciplines is through post-coordinate indexing. When used in conjunction with systems of faceted classification it becomes possible to create new knowledge implicit in studies that have already been published but from which no inferences have been drawn. This new knowledge is created by generating new combinations of concepts, thereby giving information systems a role in the stimulation of creativity.

Classification theorists as early as Ranganathan had recommended the use of classification for creating new knowledge by using vacant class numbers to suggest the creation of corresponding new subjects. Ranganathan, and later Jolley, hoped that a self-perpetuating classification that assigned to logical places subjects that did not exist when the scheme was planned would be capable of revealing gaps in our knowledge.

Zwicky developed a technique which he considered to be a new method of classification known as morphological analysis. Using a combinatorial approach, all possible solutions to a given process are generated, then evaluated to determine which one is the most suitable. Morphological analysis requires storing and manipulating various tables, whose creation could be greatly facilitated by using classification schemes and thesauri to identify relevant categories.

Classification could also play a role in finding unknown connections in the literature. Farradane suggested developing techniques based on a form of indexing that is suitable for making inferences. An indexing system that recognizes logical relationships would offer great potential in creating new knowledge. Swanson has proposed using a systematic trial and error search strategy. This consists of retrieving a set of references on a topic and scanning subtitles for words or phrases that might suggest links. These words or phrases are then used as search terms to retrieve documents, which are then scanned to determine whether there are any concepts that are linked to the original topic in a logical way. If none are found, the idea of the logical connection is assumed to be original.

A major requirement for systems supporting creativity will be the development of improved means of representating the information needed to bring out "hidden" relationships, patterns, and analogies. One way to accomplish this might be to develop relational indexing. The other would be to utilize techniques of knowledge representation used in expert systems. One of these methods might provide a means of finding undiscovered public knowledge that does not depend on the interests or ability of the searcher. Ultimately it could result in a system that traces relational paths of associations, to provide the user with previously unseen connections and associations that could result in new discoveries and new subject areas (10, p.298).

5. Theoretical Approaches to a Universal Classification Scheme

Along with the role of automation in classification, theorists are working on a number of other vital issues. One subject has perhaps engaged them more than any other. That is the effort to develop an absolute, general, universally valid classification scheme.

Researchers working on this problem agree that it poses a number of obstacles to its solution. For example, disciplinary main classes tend to have the effect of freezing the structure of knowledge. Yet knowledge itself is constantly changing as new discoveries are made, and old ones replaced.

Not only are the boundaries of knowledge expanding, the relationships between different areas of knowledge are constantly in flux. The growing importance of interdisciplinary subjects is difficult to accommodate in a scheme based on

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

disciplines. Disciplinary classes thus lack an absolute basis which would keep them universally valid whatever changes occur in the structure of knowledge (11, p.109).

There are numerous social forces that also present difficulties to the development of a universal system. Classification schemes and their categories, divisions, and subdivisions, are based on social consensus about knowledge. Yet consensus itself differs from one society to another, one historical period to another, as well as by discipline.

Given this social basis of classification, it becomes difficult to create a universal classification system that is free of nationalistic or ideological biases. Often in such a system, socially acceptable concepts are given prominence in a hierarchy, while socially unacceptable ideas or terms are not. What was intended to be a universal scheme thus turns out to be a socially stratified hierarchy of knowledge that is permeated by ideological bias. Classifications are mirrors that reflect their time, place, and society.

Imagine if the UDC had been constructed by working class women from Third World countries. It might take on a very different character than its present form. In its current version, it cannot help but reflect the perspective of American and European middle class white males. Not only the hierarchy of concepts, but the very choice of which subjects the classification seeks to enumerate is socially influenced. The Broad System of Ordering, concept-based systems like the Information Coding Classification (12) and new technologies such as computerized switching languages may present alternatives for minimizing the inherent biases of "universal" systems (13, p.396).

6. Concept Theory as an Approach to Universal Classification

In an attempt to solve the problem of fixed structures in universal classification systems, Ingetraut Dahlberg has proposed a new theory of classification based on concepts and definitions. It does not have a classificatory frame or structure to hamper periodic efforts to revise or update the system. The theory assumes that knowledge is social and verifiable and thus in need of regular updating.

The theory is based on the idea that classification does not, as is commonly assumed, deal with objects or terms, but with knowledge. This includes both knowledge about items and the organization of that knowledge.

Knowledge can only be generated by statements about something. This "about something" or item of reference, she calls the *referent*. Whenever a word or term is used to designate something about which a statement has been made, it is called a *designation*. A concept is defined as a *unit of knowledge* that comprises necessary and verifiable statements about a referent, and is represented by a designation.

There are four kinds of formal relationships between concepts. These she calls identity, inclusion, intersection, and exclusion. Besides formal relationships, there are also contents related or material relationships between them. These include the generic relationship, the partition, the opposition, and the functional relationship.

The generic relationship between concepts is the relationship of a broader or narrower concept. This relationship builds up conceptual hierarchies, often represented as a tree. The partition relationship occurs when the whole is split up into itself and the concept of its parts. The opposition or complementary relationship occurs in concepts which include the possibility of a positive and a negative kind of characteristic or an opposite or complementary one. The functional relationship consists of two concepts that find themselves syntagmatically related, such as a subject and its predicate or predicate and its complements.

Dahlberg says that knowledge elements derived from statements about referents become the components of concepts which are called characteristics. There are many different possible kinds of characteristics. The broadest kinds of characteristics are called Categories and their Subcategories, they have also been called *Form Categories*.

In contrast, there are also *Categories of Being* derived from Aristotle just as the Form Categories, viz. inanimate, animate, mental, and divine being. Dahlberg extends these to nine levels using the integrative level theory of J.K. Feibleman and Nicolai Hartmann. These nine levels are 1) Being of structure and form, 2) Being of matter and energy, 3) Being of cosmos and earth, 4) Biological being (plants and animals), 5) Anthropological being (mankind), 6) Social being (society), 7) Material Products (material artefacts), 8) Intellectual Products (knowledge, information), and 9) Spiritual Products (spiritual artefacts, language, music, art). Any combinations of this kind with form categories (such as objects, properties, activities) generate subject categories, which themselves may serve as the starting points for the formation of subject groups and fields (14, 15).

7. Concept Systems as Definition Systems

Concept systems can be utilized as definition systems. A definition can help in classification by explaining the contents of a concept. Defining a concept as broadly as possible yields a generic definition. By defining a concept as a component of an object, one obtains a partition definition, whose structure follows the partition relationship. Similarly, if a concept is defined as a negation or any kind of opposition, then the symbol representing it is the opposition definition. Finally, the functional definition provides a concept structure which not only comprises the totality of elements and characteristics, but also maps their syntactical relationships. The functional definition is also used to define disciplines and subject fields.

Dahlberg believes her concept and definition theory has applications for the field of classification. This is because it demonstrates the relation between concepts and knowledge, serves in the construction and reconstruction of concepts, helps analyze concepts according to characteristics, facilitates the comparison and correlation of concepts, explains conceptual relationships, categorizes concepts and characteristics of concepts and helps clarify the structure of concepts, among other reasons (15).

One investigator, P. Rolland-Thomas, claims that until recently, most attempts to construct broad based general

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

classification schemes have used the sciences as the model for all knowledge. The worldwide emphasis on scientific research has created a strong demand for access to scientific documents. Thomas points out that recent investigators attempting to devise universal encyclopedic classification systems have realized that any such efforts are invalid without also including the arts and humanities. They have concluded that a theory of knowledge needs to be created for the humanities (16) and much attention has recently been focused on that task.

8. Classification Theory in the Humanities

Ironically, although Rolland-Thomas doesn't mention it in her article, the development of *The Art and Architecture Thesaurus* had begun several years prior to its appearance. The AAT is still under development, and as of 1990 consisted of nearly 40,000 terms representing the field of art and architecture.

The thesaurus was created to provide a link between the object, its images, and related bibliographic material. It was developed because of general dissatisfaction among art librarians with the coverage LCSH gave to the field of art and architecture. Focusing on Western art and architecture, it builds upon vocabulary already in use in the field.

It was originally designed to use a simple alphabetic listing of hierarchies. Later, it was revised into a faceted classification scheme. It starts with the most abstract concepts and proceeds to hierarchies containing terminology for styles and periods of art, agents, activities, materials, and then object types.

The AAT tries to assimilate both the language of scholars in the field and the more popular language found in basic literary sources. This reflects the view of its creators that while it is desirable to have a comprehensive, standardized vocabulary, a thesaurus cannot be stagnant or authoritarian. They envision the thesaurus "as a living tool; a body of language that can be added to and changed as it responds to the needs of its users" (17, p.653).

One of the difficulties of this undertaking is that there is some difference of opinion regarding what constitutes the humanities. History is a discipline that is claimed by both the humanities and the social sciences. Another obstacle is the form that primary works take, such as a painting, a music score, or a work of fiction. Of all these, fiction is closest to the documents for which techniques of subject analysis have already been developed.

9. A Theoretical Approach to Classifying Fiction

Beghtol believes that adequate classification systems have not been developed to access the content elements of primary works of fiction. As a result, they are often classified by creator rather than by subject matter. Since more detailed content access would prove useful, efforts have been made to devise such a scheme (18).

Extracting data from fiction presents special difficulties for the classifier. The presence of all sorts of fictional entities in the form of unknown or unnamed creatures is possible. Fiction may also blend together "real" and "unreal" elements, which may magically transform into one another without reason. Characters may change gender, occupation, or location in the course of a story.

Fiction also contains a good deal of ambiguity. The actions of characters, their motives, and sometimes the relations between one part of the plot and another, are not always clear. The question becomes one of whether this kind of data can be classified, and if so, whether different catalogers would agree on subject headings.

One way a classification theory can handle this ambiguity is through fuzzy sets. Fuzzy set theory assumes there are no strict well defined categories in the real world. Instead of simply assigning an item to a category, a fuzzy set acknowledges its ambiguity by assigning it a degree of membership in a certain class, signified by values between 0 and 1.

Another way to handle the uncertainty of fiction is to introduce a notational element expressing ambiguity. This notational element could be used to categorize documents demonstrating ambiguity at any or all available depth of specification. A default category "Other" could also be used for problem data that will not fit any other category. It is worth noting that whatever theories are developed to cope with the difficulties of classifying fictional data may also prove quite useful for classifying the content of other kinds of documents (19, p.47).

From the preceding sample it can be seen that classification theory is a lively and rich area of exploration. A number of issues remain unresolved and many of the problems are rather daunting, but this is a part of what makes it so intriguing and important. Yet despite all the interesting work being done internationally, the question still remains as to why America is so underrepresented in these efforts.

10. Discussion

In order to try to account for the paucity of classification theory by Americans in this review, the current literature on librarianship was carefully searched for clues. Only one article was found that mentioned this phenomenon, which itself seemed odd. Apparently no one seems to have taken note of the situation. Or if they have, they have not bothered to write about it. Instead, American cataloging journals are filled with endless numbers of articles discussing the most idiosyncratic details of descriptive and subject cataloging, with hardly a thought given to any broader perspective.

The lone article that does mention this unfortunate state of affairs was written five years ago by Richard Halsey, then dean of the School of Information Science and Policy at SUNY Albany (20, p.93). Halsey speculates that classification theory has not been given the attention it deserves in the United States for several reasons. These reasons are described very loosely, without being too specific or offering much in the way of examples or evidence, which tends to underscore the tentative nature of his argument.

According to Halsey, the study and creation of taxonomies and classification systems requires a command of language and culture that Americans lack. He does not specify any

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

particular language, nor why it is necessary to have mastered it. Nor does he define what he means by "culture," or why it is important to classification design. One clue though, is offered when he confesses "Jade strokers and speculative thinkers are a minority of our most educated population."

He seems to attribute this lack of speculative thinking to a decline in education and literacy, which itself has several causes. Among these are the diminishing literacy caused by competition from the media and mass culture. Perhaps it is no coincidence that Halsey was writing this at the time when Alan Bloom and Ed Hirsch were generating considerable controversy on campus with their books *The Closing of the American Mind* and *Cultural Literacy*.

Bloom chronicles the decline of interest in reading the *Great Books* and classics which he believes is brought about in part by competition from the mass media. He also blames the narrow disciplinary fragmentation that has occurred in many fields, led by the sciences, and the corresponding loss of perspective this has created due to specialized vocational training. Bloom suggests that the sciences and other practical disciplines claim to be metaphysically neutral and thus have no need to ask the big questions anymore: "The kinds of questions children ask: Is there a God? Is there freedom? Is there punishment for evil deeds? Is there certain knowledge? What is a good society? These were also the questions addressed by science and philosophy. But now the grown-ups are too busy at work, and the children are left in a day care center called the humanities, in which discussions have no echo in the adult world." (21, p.372)

He contrasts this situation with the one in Europe where "school children are taught philosophy, and it seems to be something real." The only philosophical movement that America has given birth to is pragmatism, and America is an inherently pragmatic culture, concerned with the observable and the measurable.

If American school children lack much of a background in philosophy, American library students lack much in the way of exposure to classification theory. This is a result of an American tradition in library education that began with its founders. Dewey himself may have constructed an elaborate classification, but he did not encourage the students in his library school to do so, and neither did most of his successors.

This heavy emphasis on vocational education at the expense of theory reflects a more general American emphasis on the utilitarian value of knowledge. Practical knowledge benefits the whole society and is therefore democratic. Philosophy, however, is a luxury enjoyed by an intellectual aristocracy and is therefore elitist.

The emphasis that American library education placed on vocationalism struck the eminent library educator Pierce Butler as being rather ironic. He saw library education as rooted in epistemology, the branch of philosophy that studies the nature and extent of knowledge. If librarianship is concerned with the management of knowledge it is conceivably the most interdisciplinary of all disciplines. Similarly, if it is concerned with the philosophy of knowledge, it is potentially the most deeply philosophical of all professions (22, p.176).

If one is hard pressed to find much theory in American library schools, the same holds true for cataloging courses. A recent examination of some of the better known American texts such as Wynar, Chan, and Immroth, reveals just how little theory students are exposed to. Wynar devotes less than twenty pages to a general discussion of classification, followed by 270 pages covering schedule formats, number building, Cutter numbers, subject headings, automated indexing, authority files, online bibliographic networks, cataloging routines, and filing. Throughout, the emphasis is on technique, not theory (23).

Chan's focus is even more narrow, being almost exclusively concerned with LC classification. There is a brief historic introduction, followed by mostly technical discussions of Cutter tables, Cutter numbers, tables, individual classes, classifying serials and collections, corporate headings and similar topics (24).

There is a general absence of theory, and the whole approach is similar to Wynar's practical cookbook style presentation.

To get a sense of how the education of American catalogers compares with that of students abroad, some of the better known texts used in England were also examined. Jack Mills' text on library classification has a brief theoretical discussion in the opening two chapters of the book, and a final short chapter that is somewhat theoretical as well. But overall, his treatment of classification is practical and applied, with the emphasis being heavily on technique. He is a sort of English Bohdan Wynar, at least as evidenced in this text (25).

A. C. Foskett's text on the subject approach to information is altogether different. From the very first chapter "Theory of Information Retrieval Systems," the approach is strongly theoretical and conceptual. Along with the theoretical approach, there is also much more emphasis given to automation, indexing, and evaluation than in the American texts (26). The whole concern is with understanding classification concepts rather than applying them. The discussion is thoughtful and sophisticated, with the emphasis on explanation rather than demonstration.

Two texts by Derek Langridge are even more theoretical in their approach. His *Approach To Classification* (27) reads less like a text than a book of aphorisms. It consists of five parts, each of which contains a brief statement of classification followed by single page explanations of their significance. The directness and simplicity and brevity is reminiscent of Strunk and White's *The Elements of Style*. Classification is treated not so much as a technique than as a subject, and the book is refreshingly free of jargon. At one point, Langridge says candidly: "*Classification is sometimes discussed as if it were solely a technique for arranging books on the shelves of libraries. In fact there is a school of thought, predominant in USA, that does attempt to restrict classification to this role, but this is a mistaken view.*" Langridge holds a much broader philosophical view of classification as the fundamental human activity for making sense out of the world that permeates all aspects of life. His other text *Subject Analysis: Principles and Procedures* (28) has a more specific focus, but its approach is also highly theoretical and conceptual in nature.

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

From this comparison of American and English cataloging texts, it is evident that students abroad receive much more exposure to classification theory in their training than students in the U.S. This suggests a strong reason why theory is so insignificant on the American scene. American catalogers who have had little exposure to it can hardly be expected to have any understanding or appreciation of its value to classification.

This raises an important question. What is the value of theory to classification and cataloging? Very little seems to have been written about the role of theory in cataloging and classification, or in librarianship in general. Those who have written about it however, tend to agree that it serves a vital function.

One important function of theory is to establish an agenda for research. Theory gives researchers an idea of the extent of what is known by synthesizing it. Equally important, by identifying gaps, it suggests what remains to be investigated. Theories are created provisionally, with the understanding that subsequent research will either support or refute them. Theories also supply a rationale for, or an argument against, current practices in the field (29, p.358).

Theory can increase understanding and guide practice. It may also be useful in a more fundamental sense of getting one to think about a problem. Theories can serve as a prelude to a more systematic examination of the topic (30, p.17). They can put things in perspective, or provide a new and different perspective.

Another purpose of theory is to spur innovation. It can help generate the production of new ideas. By causing the questioning of existing practices, established traditions, and unquestioned assumptions, it can lead to better ways of doing things (31, p.153).

The practice of any problem solving activity usually proceeds in the most obvious and expedient manner. Problems are thus often treated on a superficial level. Theory allows for deeper analysis and deeper insight. It invariably leads one to ask "why do I do these tasks?" instead of "how do I do these tasks?"

One study has suggested that the notion that theory and practice are separate is actually a myth. They are really different aspects of the same thing. Every theory is related to practice in some way. And every practice is ultimately based on some kind of theory. It is simply a matter of strengthening the connection (32, p.28).

Theory enables one to explain relationships among phenomena. It allows one to generalize beyond one's particular situation to other incidents or cases to better understand them. By doing so, it goes beyond explanation to open up the possibility of prediction. To the extent that one can predict, one may be able to control a situation to some degree (33. p.228).

Everyone is a "theorist" to the extent that he/she holds beliefs about something. The value of theory is that it makes explicit what we all do implicitly. By making one's beliefs explicit, one becomes more aware of one's ideology and its strengths and weaknesses. Becoming conscious of the assumptions and principles that underlie one's techniques, operations, and services is the first step toward improving them (34, p.3).

Theory may be a determining factor in whether classification and cataloging in America (and, for that matter, librarianship) deteriorates into mere craft or develops as a profession. In order to justify their claims to be a profession, librarians must master a body of knowledge as well as techniques, because a profession consists of both. Theory in particular and research in general provides the knowledge base which characterizes a profession (35, p.375).

Being a professional involves prolonged training in a body of abstract knowledge which is conceptual, not just technical. Professionals not only possess this knowledge, they help create it. Librarians thus need a stronger theoretical underpinning for their actions. Exposure to day to day procedures and techniques based on rulebooks or manuals without reference to more general theoretical principles does not properly constitute a profession (36).

Theory can thus be seen to play an important role in the practice of any activity. The estrangement of American classification from its theoretical, philosophical, and conceptual foundations cannot continue without deleterious consequences. Until this situation is corrected, it is a discipline destined to drift aimlessly, without a long term vision and with limited prospects for development or renewal (37, p.9).

As one study of the library field warned over a decade ago, a discipline that does not formulate theory does not control its own scientific or technical advance (38, p.391). If American classification continues to function as more of a service than a science, it runs the risk of forfeiting control of its destiny. For example, the increasing technical feasibility of the electronic production of journals and books, including direct dissemination to clients, necessitates that librarians take a leadership role in theory and research in this area. Failure to do so may result in other fields taking the lead, with the result that the library community might end up as an occupational dinosaur.

In all fairness to the American classification community, if they are to be expected to take theory more seriously, not only the quantity of theory but the way that theory is disseminated must change. The present situation in America is one in which infrequent articles on classification theory appear in obscure journals or conference reports. Since most catalogers do not read these publications, the impact of the articles on working catalogers can be at most indirect. They may eventually result in practical applications if some cataloger happens to come across one by chance, but the few theory articles that appear should be published in more of the mainstream journals so that they are able to reach a wider audience (39, p.2).

There is also room for improvement in the quality of classification theory currently being generated both here and abroad. Presently, it tends to lack cohesiveness and synthesis. Articles on theory tend to be random and noncumulative (40, p.487). There appears to be little collaboration between theorists in attempting to solve longstanding problems. Theorists pay little attention to each other's work, and essentially

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

function as independent auteurs. Theorists are even inconsistent with their own individual research, in that their theoretical works often do not build on their previous theoretical efforts. They frequently tend to generate new theory on new topics without first bothering to revise or refine previous theories they have published.

This is in no way to diminish the importance of theory, only to note that there is much room for improvement. More and better theory must be created, because without theory, the American classification and cataloging community will have no place to turn for its new ideas.

Acknowledgment

The author wishes to thank Professor Pauline A.Cochrane for her guidance in developing this article.

References

(1) Neill, S.D.: The Dilemma of the Subjective in Information Organization and Retrieval. J.Doc. 43(1987) p.204.
(2) Belkin, N.J., Oddy, R.N., Brooks, H.M.: ASK for Information Retrieval: Pt.1. Background and Theory. J. Doc. 38(1982)2, p.63.
(3) Frohmann, B.: Rules of Indexing: A Critique of Mentalism in Information Retrieval Theory. J. Doc. 46 (1990) p.94.
(4) Grolier, E.de: In Search of an Objective Basis for the Organization of Knowledge. In: Ordering Systems for Global Information Networks. Proc. 3rd Int. StudyConf. on Classification Research, Bombay, India, Jan. 6-11, 1975 Bangalore: FID/CR and Sarada Ranganathan Endowment for Library Science 1979. p.64.
(5) Jones, K.P.: Towards a Theory of Indexing. J.Doc. 32(1976) p.121.(6) Fugmann R.: On the Practice of Indexing and its Theoretical Foundations. Int. Classif. 7(1980) p.13.(7) Fugmann, R.: Indexing Quality: Predictability versus Consistency. Int. Classif. 19 (1992) 1, p.21.
(8) Fugmann, R.: Unused Possibilities in Indexing and Classification. In: Tools for Knowledge Organization and the Human Interface. Proc. 1st Int. ISKO Conf., Darmstadt, 14-17 Aug. 1990. Frankfurt: INDEKS Verlag 1990. p.65.
(9) Veltman, K.: Computers and a New Philosophy of Knowledge. Int.Classif. 18(1991)1, p.11.
(10) Davies, R.: The Creation of New Knowledge by Information Retrieval and Classification. J. Doc. 45(1989) p.298.
(11) Buchanan, B.: Theory of Library Classification London: C. Bingley 1979. p.109.(12) Dahlberg, I.: ICC - Information Coding Classification - Principles, Structure and Application Possibilities. Int.Classif. 9(1982)2, p.87-93
(13) Wilson, A.: The Hierarchy of Belief: Ideological Tendentiousness in Universal Classification. In: Williamson, N.J., Hudon, M. (Eds.): Classification Research for Knowledge Representation and Organization. Proc. 5th Int. Study Conf. on Classif. Research, Toronto, Ont., 1991 Amsterdam: Elsevier 1992. p.396.
(14) Dahlberg, I.: The Basis of a New Universal Classification System seen from a Philosophy of Science Point of View. In: (same source as) (13), p.187-197
(15) Dahlberg I.: Concept and Definition Theory. In: Classification Theory in the Computer Age: Conversations Across the Disciplines. Proc. Conf. Nov. 18-19, 1988 Albany, NY. Albany, NY: Rockefeller College Press 1989. p.23.
(16) Rolland-Thomas, P.: Towards the Establishment of the Validity of Encyclopedic Library Classification Systems. In: Universal Classification: Subject Analysis and Ordering Systems. Proc. 4th Int. StudyConf. on Classif.Research, Augsburg, June 28-July 2, 1982, p.49.
(17) Petersen, T: Developing a New Thesaurus for Art and Architecture. Library Trends 38(1990)4, p.653.(18) Beghtol, C.: Access to Fiction: A Problem in Classification Theory and Practice. Pt. I. Int. Classif. 16(1989)3, p.134.
(19) Beghtol, C.: Toward a Theory of Fiction Analysis for Information Storage and Retrieval. In: (same title as) (13) p.47.
(20) Halsey, R.S.: Implications of Classification Theory in the Computer Age for Educators of Librarians and Information Professionals. In: (same source as) (15) p.93.
(21) Bloom, A.: The Closing of the American Mind. New York: Simon & Schuster 1987. p.372.
(22) Shera, J.H.: Libraries and the Organization of Knowledge. Hamden, CT: Archon Books 1965. p.176.
(23) Wynar, B.S., Taylor, A.G.: Introduction to Cataloging and Classification 8th Ed. Englewood, CO: Libraries Unlimited 1992.
(24) Chan, L.M.: Immroth's Guide to the Library of Congress Classification 3rd Ed. Littleton, CO: Libraries Unlimited 1980.
(25) Mills, J.: A Modern Outline of Library Classification 5th Ed. London: Chapman & Hall 1967.
(26) Foskett, A.C.: The Subject Approach to Information. 4th Ed. London: C. Bingley 1982.
(27) Langridge, D.: Approach to Classification for Students of Librarianship. London: C. Bingley 1973.
(28) Langridge, D.W.: Subject Analysis: Principles and Procedures. London: Bowker-Saur 1989.
(29) Schmidt, Nancy J.: Research without Theory: Data Collection as an End in Itself (Weakness of Library Research). J.Acad. Librarianship 17(1992) p.358.
(30) Neill, S.D.: Can there be a Theory of Reference? Reference Librn. 18(1987) p.17.
(31) Intner, Sh.S.: Theory and Practice or Theory versus Practice: Fundamental Issues and Questions. In: Intner, Sh.S., Vandergrift, K.E. (Eds.): Library Education and Leadership. Metuchen, NJ: Scarecrow Press 1990. p.153.
(32) Ruhig Du Mont, R.: Bridging the Gap between Theory and Practice: The Role of Research in Librarianship. Training & Education 6(1989) p.28
(33) Grover, R., Glazier, J.: A Conceptual Framework for Theory Building in Library and Information Science. Libr. & Inform. Sci. Rev. 8(1986) p.228.
(34) Busha, Ch.H.: The Meaning and Value of Theory. Drexel Libr. Quart. 19(1983) p.3.
(35) Lynch, M.J.: Research and Librarianship: An Uneasy Connection. Libr. Trends 32(1984) p.375.
(36) Goode, W.J.: The Librarian: From Occupation to Profession? Libr. Quart. 31(1961)
(37) Studwell, W.E.: Perfidious Policies: Or, Lack of Theory Equals Lack of Firm and Constant Direction for LC Subject Headings. Technicalities 10(1990) p.9.
(38) Heilprin, L.P.: The Library Community at a Technological and Philosophical Crossroads: Necessary and Sufficient Conditions for Survival. J. Amer. Soc. Inform. Sci. 31(1980) p.391.
(39) Katz, B.: The Influence of Theory and Research in the Practice of Reference. Reference Librn. 18(1987) p.2.
(40) Howard, H.: Organization Theory and its Applications to Research in Librarianship. Libr. Trends 32(1984) p.487.

Brian Quinn, 83, Violet Ave. Floral Park, NY, 11001, USA.

https://doi.org/10.5771/0943-7444-1994-3-140 https://www.inlibra.com/de/page/agb - Open Access -