Is Classification Theory Possible? Rethinking Classification Research

1. Introduction

Theoretical context independent explanations of classification could enhance the universality of classification research and make knowledge about classification available to settings other than traditional libraries. There is a tremendous need for constructing classificatory structures in a range of settings many of which are far removed from the environment in which classification theory and research has been practiced in the last century and a half. The construction of classificatory structures on the Internet, intranets, and in knowledge management systems has received some attention lately. The question examined here is whether it is possible to create a single theory of classification that applies to the range of contexts in which classificatory structures are applied.

The object of this paper is to question the assumption that bibliographic classification theory can resemble scientific theories. It is argued that the context of any classification influences the use and understanding of the classification to such a degree that the classification cannot be understood separate from its context. Furthermore, the development from being a novice classifier or classificationist to becoming an expert is explored. It is assumed scientific theories must relate as much to the activity of novices as to the activity of experts and that scientific theories explain both what it is that novices do and what experts do. It is argued that expertise is achieved not through a correct application of a classification theory but through experiences and adjustment to a particular context and that the activities of novices are quite distinct from the activities of experts in that experts draws on the context of the situation and that novices do not.

2. Theory of Classification

Langridge (1976) provides an account of the principles of constructing knowledge organization systems and the theoretical underpinnings of different approaches. He identifies four principles that have guided construction of knowledge organization systems: 1) ideological, 2) social purpose, 3) scientific, and 4) the disciplines. The ideological principle organizes knowledge according to an ideology that the knowledge organization system serves. Langridge gives the examples of "the Christian schemes of the Middle Ages and the Soviet scheme which substitutes for the Bible and Christianity the works of Marx and Lenin and the 'religion' of communism" (Langridge, 1976, p. 4-5).

As an example of a knowledge organization system guided by social purpose, Langridge refers to Ranganathan's example of an ancient Indian classification that organizes knowledge "in an order of decreasing immediacy of
social use" (Langridge, 1976, p. 5). The system only has four classes and has not been used to organize bibliographic material.

The last two principles--scientific and disciplines--are the only two principles that have been used in modern classification theory. The latter is Langridge's contribution and the former the most widespread.

Langridge's idea is that is a set of disciplines of knowledge that most philosophers agree on and that these disciplines could be used as the basis of classification. He surveys seven philosophers' (Mortimer Adler, Ernst Cassirer, R.G. Collingwood, Paul Hirst, Charles Hillis Kaiser, Philip H. Phenix, and Louis Arnaud Reid) categorizations of the sciences and finds a similarity between them. Each of these philosophers divides the entire universe of knowledge into 4-7 fundamental disciplines of knowledge. They do not sub-divide the disciplines. Langridge argues that a general classification scheme should use these fundamental disciplines of knowledge to organize the universe of knowledge and then further subdivide the individual disciplines by the phenomena studied or methods used. Langridge notes that it more common to divide according to specialization but that "today it seems likely that it is more helpful to divide by phenomena studied" (Langridge, 1976, p. 9).

Common for these three principles is that they use categorizations that already exist as the basis for developing a bibliographic classification. This basis could either be an ideological categorization, a social categorization, or a philosophical categorization. To make a successful bibliographic classification based on any of these principles it is required that the categorization principle is widely accepted and recognized.

Langridge names the fourth principle the scientific principle. It is probably the principle that has received the most attention in modern classification theory. The principle began with the work by Richardson (1935, first edition 1901) whose view can be summed up by "two statements: the order of the sciences is the order of things, and the order of the things is the order of their complexity" (Langridge, 1976, p. 5). Although the theory applies to natural objects it has not always been made clear, that the theory's relation to man made things was a matter for investigation and should not be taken for granted. Langridge finds that the principle at first sight meets his requirement of neutrality but that "serious objections ... arise from [the] underlying philosophical assumptions" (Langridge, 1976, p. 5). These assumptions are: 1) Science is the paradigm of all knowledge, 2) Knowledge is a simple unity, and 3) There is an objective external world. Langridge dismisses the scientific principle because it is tied to these positivistic assumptions.

3. Positivistic Epistemology

Harris (1986a; 1986b) has investigated the historical roots of the commitment to positivism in LIS. He finds that it can be traced back to the Carnegie Cooperation's establishment of the Graduate School of Library Science at the University of Chicago in 1928. Harris argues that the first faculty at the school defined and established what he at one point calls "a 'psychosociological' research program" (Harris, 1986a, p. 517) and at another point calls "a positivistic epistemology" (Harris, 1986b, p. 218). The school further succeeded in forming a conception in the LIS field over the next several generations that 'good research' is founded in such an approach (Harris, 1986b, p. 218). Harris further argues that the
belief in the possibility and desirability of a science of librarianship recalls the situation of the social sciences in the 1950s (Harris, 1986a, p. 528; 1986b, p. 220-221) summarizes the case against positivism in the social sciences in three points,

1 the social sciences have never been able to generate "scientific paradigms" that might govern research in ways analogous to the hard sciences;
2 the social sciences simply cannot sustain the essential division between the subject and the object of research so central to the positivist epistemology;
3 the value free pretensions of the social sciences have been proven to be mystifications designed to camouflage the extent to which the social scientist is governed by prejudgments and domain assumptions.

LIS must separate itself from a positivistic epistemology in the same way that it has been necessary in the social sciences, even though a positivistic or quasi-positivistic approach remains the most widespread approach to research and thinking in the LIS field. This is seen by the emphasis on objectivity, generalization of findings, empirical research, quantification, etc. Most importantly such an approach to the "field has severely limited the range of questions that can be investigated and has rigidly defined the characteristics of relevant answers" (Harris, 1986b, pp. 221-222).

An ideal theory of classification that follow the neutral, objective and positivistic line of thought should be able to prescribe how a set of documents should be organized and predict the consequences of the organization. The theory should furthermore apply to all kinds of different settings, users, and document types. Flyvbjerg (2001, pp. 38-39) sums up the requirements of an ideal theory based on his readings of Hubert Dreyfus and Pierre Bourdieu. Flyvbjerg finds that ideal theories have six characteristics:

• Explicit. A theory needs to be laid out in such detail that any reasoning human being is able to understand it. The theory must not stand or fall on interpretation or intuition.
• Universal. The theory must apply at all times and in all places.
• Abstract. The theory must not require reference to concrete examples.
• Discrete. The theory must be formulated with context-independent elements; it cannot refer to human interests, traditions, institutions, etc.
• Systematic. The theory must constitute a whole in which the context-independent elements are related by laws or rules.
• Complete and predictive. The theory must be complete in the sense that it covers its whole domain and it must be predictive in the sense that the theory must specify the effects of the elements.

These characteristics cannot, of course, be fully realized in all situations. It is only possible to approach them to varying degrees. However, they do characterize what a theory is and what we can expect from a theory. These characteristics are the frame of reference for understanding the scope and possibilities of classification theory and insofar as a general classification theory should be regarded as a scientific theory about classification of documents, it needs to conform to these characteristics of scientific theories.
4. Context, Theory, and Classification

Everyday practical librarian and information specialist work is based on adaptability to individual work domains, understanding of individual needs, interpretation of documents for representation, etc. all of which takes place in particular contexts. Much research in classification, on the other hand, is based on an assumption that classification research should be objective, neutral, and replicable.

In a personal statement in one of the two 50th anniversary issues of JASIS Michael Buckland talks about two traditions within library and information science, namely the computational tradition and the document tradition. He defines the document tradition, as being concerned "with knowledge, meaning, learning, description, language and ambiguity, therefore, any view of it remains incomplete unless some roots in cultural studies, in the humanities and qualitative social sciences, is acknowledged" (Buckland, 1999, p. 971). It is very difficult to discuss classification of documents without having focus on the meaning of the documents and the language they use. Any investigation into classification must start with a clear and explicit theory of language, meaning, and influence of context on that. Furthermore, any empirical investigation of classification needs to explicitly take into account the context of situation studied. Therefore, classification research must in reality be studies of human activity, language, and interpretation.

Any application of an overreaching bibliographic classification theory will in such practical circumstances fall short. The reason for such a conclusion is that a theory that makes explanation and prediction possible necessarily must exclude the concrete context of everyday activity. However, this exclusion of context makes explanation and prediction impossible.

5. Novice and expert classifiers and classificationists

Classification is primarily a human activity and research and studies of classification is an attempt to understand human activity. One of the ways to understand the limitations of a general classification theory is to see how people learn how to classify. Dreyfus and Dreyfus (1986) have identified five stages of development from being a novice at some task to becoming an expert. People acquiring a certain skill will pass through these stages in becoming better and better at the skill and mastering it in the end. People at one stage will be better at the skill than people at a previous stage. There are virtually no limits to the types of skills that Dreyfus and Dreyfus' stages cover, as long as the skill can be acquire by any rational human being (as for instance bicycling, playing chess, surgery) although not all people will be able to reach the higher stages in all skills.

The five stages and some simple notions of them are:

Stage 1. Novice. "The novice learns to recognize various objective facts and features relevant to the skill and acquires rules for determining actions based upon those facts and features" (Dreyfus and Dreyfus, 1986, p. 21). The novice will only be able to apply the learned rules in contexts and situations with which the novice is familiar. The learned rules are therefore context free rules. The rules are context free in the sense that the novice has learned the rules objectively and will apply the rules to the particular situations with no regard for the situation itself. The novice will apply the learned rules in situations that match the learned proto-situation and
will not be able to accommodate for the given context of a new and different situation.

Stage 2. Advanced beginner. As the novice becomes familiar with the skill, she advances to a marginally acceptable level of performance. This is done "through practical experience in concrete situations with meaningful elements, which neither the instructor nor the learner can define in terms of objectively recognizable context-free features" (Dreyfus and Dreyfus, 1986, p. 22). The advanced beginner begins to recognize and apply previously experienced situations to new situations. Hence the advanced beginner uses both context-free rules and "situational" elements in performing a task.

Stage 3. Competence. Whereas the novice and the advanced beginner merely followed a set of rules in coping with a situation, the competent performer sees a situation as a set of facts. The competent performer approaches the situation with a plan to organize the situation in mind. The situation is then examined according to the plan, and only those factors which are important to the given plan are given consideration. Since the competent performer acts according to a plan which she has chosen herself, she "feels responsible for, and thus emotionally involved in the product of choice" (Dreyfus and Dreyfus, 1986, p. 26). The novice and the advanced beginner, on the other hand, act according to specified elements or rules, and thus feel that an unfortunate outcome is the result of inadequately specified elements or rules.

Stage 4. Proficiency. Unlike the previous performers, the proficient performer does not rely on rules in performing a task. At this stage she freely uses her intuition. This should not be confused with either irrational conformity or guessing. To guess is to reach a conclusion when one does not have sufficient knowledge or experience. Intuition here means the "sort of ability we all use all the time as we go about our everyday tasks" (Dreyfus and Dreyfus, 1986, p. 29). The proficient performer is characterized by her ability almost immediately without conscious effort to follow a plan after encountering a problematic situation. Yet it is unlikely that the proficient performer will be able to explain in detail how she solved the situation.

Stage 5. Expertise. "An expert generally knows what to do based on mature and practiced understanding" (Dreyfus and Dreyfus, 1986, p. 30). The expert does not see problems in some detached way. She is so much a part of the situation that she does not need to be aware of the situation to handle it. "When things are proceeding normally, experts don't solve problems and don't make decisions; they do what normally works" (Dreyfus and Dreyfus, 1986, p. 31). When things work as they normally do, the expert is able to deliberate before acting. But this deliberation is not based on calculative problem solving. Rather, it is based on critically reflecting on her intuitions. Most importantly, as the performer becomes an expert her performance becomes fluid. The expert seldom needs to consider what she does, but when she needs to she is able to do so critically. And, like the proficient performer, the expert will not be able to articulate her exact actions. The expert will likewise not be able to legitimize her decisions, but even so she will seldom make wrong decisions.

Dreyfus and Dreyfus' model contains an important jump between the three first stages and the last two stages. Somewhere around this point in the advancement from novice to expert does the person abandon using rules as the main guidance in performing the task and relies instead on intuition and context. There is,
in other words, a crucial difference in how less experienced people and more experienced people perform a certain skill. Less experienced people more or less apply a set of rules to a given situation whereas more experienced people rely on their tacit experience and intuition.

A theory of classification of documents should inform people at the first earlier stages about how to perform the task as people on the latter stages do. However, people on the latter stages perform the task without being conscious about exactly what it is that they do and they are not able to explicitly formulate what they do. Furthermore, when a person learns how to classify, the person applies some very simple context-independent rules, techniques, and theories about classification but when the person becomes more experienced, he/she will rely less on these rules, techniques, and theories and more on his/her experience, intuition, and the particular context.

6. Discussion and conclusions

The simple conclusion is that it is not possible to develop a general bibliographic scientific classification theory. The reason for this conclusion is that classification is dependent on the particular contexts to such a degree that statements about classification in reality are statements about how human beings acts and should act in particular circumstances and context. However, one of the particular requirements to a scientific theory is that it needs to be independent of particular interpretations, contexts, examples, traditions, etc. In other words, the theory must be able to prescribe how the classification should be constructed and predict the consequences of it. Furthermore, classification is a skill that only partly relies on rules, techniques, and theories; expert uses their intuition and experience when classifying. More experienced classifiers perform the task much different from less experienced classifiers. The study of experts will not directly uncover how they perform their tasks such that less experiences classify could copy the experts' way of doing things.

The upshot of this discussion is a call for rethinking classification research. Classification research is about how human beings act in particular contexts and circumstances and that is what classification research should be about. There does not and cannot exist a general universal theory about classification of documents. The best way to learn about classification is through studies of how and why people act as they do in particular contexts and circumstances. In other words, the case study method seems like an ideal method for classification research. At the center of the case study method lies the study of context-dependent knowledge and expertise.

References

