Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

Understandings of Language and Cognition: Implications for Classification Research

Hanne Albrechtsen
Birger Hjørland
The Royal School of Librarianship
Birketinget 6
DK-2300 Copenhagen S
Denmark

INTRODUCTION

In our research (e.g. Hjørland, 1992 & 1993; Hjørland & Albrechtsen, 1994, in press), we advocate a "domain-analytic" approach to Information Science (IS) based on sociology of knowledge and philosophy of science. This approach presents an alternative to the dominating individualistic views in IS, such as cognitivism. In this contribution, we shall continue our argument for more interpretative epistemologies, more emphasis on discourse analysis and a more historical orientation and draw some implications for classification research.

Besides, this contribution will discuss the difference between scientific and ordinary language and cognition. Some authors (e.g. Brier, 1992, p. 103 and Christiansen, 1994, p. 37) have interpreted and criticized "domain analysis" as implying that scientific concepts and scientific cognition should have a higher status than ordinary concepts and cognition in e.g. classification research. A typical reaction is the following citation: "Common language is the basic referent for all scientific, logical, mathematical, religious and magical languages. Scientific language is not nearer to the truth than common language" (Brier, 1992, p.103).

This reaction is in line with much contemporary epistemology. For instance, the difference found by Jacob (1994) between scientific classifications and everyday categorization constitutes two extremes, in that scientific concepts are argued to be formal, whereas everyday concepts be more flexible. In contrast Sermin & Gergen (1990) advocate that the boundaries between the two communicative domains are often blurred because the dialogue in science does not constitute a separate speech community, but draws on the same experiential basis as everyday discourse. In our opinion, such reactions make a discussion of this problem mandatory. The latter view by Sermin & Gergen questions the position of 'scientism', i.e. that scientific cognition is more true, more formal and objective than ordinary cognition, whereas Jacob's critique implies the formalism of scientism versus the flexibility in ordinary communication.

Our main thesis about the critique of "scientism" is, that the dualism between scientific and ordinary language is questionable. This dualism represents in itself a certain epistemological position. The critique of "scientism" should be a critique of certain theories and practices of science, *not* an attempt to substitute scientific concepts with lay concepts.

THE CONSTRUCTION AND USE OF SCIENTIFIC CONCEPTS, OR: HOW SCIENTIFIC IS SCIENTIFIC LANGUAGE?

Scientific concepts can, in a formalist view, be regarded as basic devices that control scientific practice via their formal definitions, or as a language domain which simply distinguishes itself

ISSN: 2324-9773

from ordinary language and everyday experience. In contrast, recent contributions exposing a more hermeneutical view of the function of language in scientific communication have found that scientific concepts can be conceived of as something much more 'fuzzy' and plastic, which cannot be pinned down by "one-word-to-one-concept" definitions. Under the headline "The Doctrine of Correct Definition", George Lakoff (1987, p. 172) presents the formalist view of scientific concepts vs. ordinary concepts:

> The metaphysical distinction between essential and contingent properties induces an epistemological distinction between two kinds of knowledge - definitional knowledge and encyclopedic knowledge. Definitional knowledge is the knowledge of the essential properties of words, and encyclopedic knowledge is knowledge of the contingent properties of words. On this [objectivist] view, the words of a language have an objective institutional status. Since words are objectively existing entities, they have essential and contingent properties. For this reason, objectivists [rationalists] hold that words have correct definitions — definitions that are objectively correct as a matter of institutional fact.

This view represents a very formal, nominalistic view of language, which regards words as labels put on some kind of objective reality, incapable of being subject to different interpretations and understandings. Further, this above distinction between scientific concepts (based on definitional knowledge) and ordinary concepts (relying on encyclopedic knowledge) presupposes, as argued by e.g. the Danish linguist Frans Gregersen (1990 pp. 125-126) for encyclopedic knowledge, that knowledge is present in meaning production and communication, and "evades historical development and variation, the two key issues for sociolinguistics" (our translation). We agree with Gregersen that knowledge and meaning production are interconnected and constantly evolving, both in "everyday, ordinary communication" as well as in scientific communication. Following Gregersen, we understand 'meaning production' in a sociolinguistic sense, i.e. that 'meaning' is not a fixed device in lexical, individual memory, but is constructed and negotiated in communication.

In the following we will introduce some recent contributions promoting more hermeneutical views of the formation of scientific concepts, conceiving them as 'fuzzy' or plastic categories, which cannot be subject to formalist definitions (Froehlich, 1994; Van Helvoort, 1993; Hohenester et al, 1988). As a case showing how standardization of scientific concepts can constrain their development, we introduce a discourse analysis of the psychiatrical classification system of mental disorders, DSM-IV (Parkinson McCarthy & Gerring, 1994). The contributions by Froehlich, Van Helvoort and Hohenstester et al show the difficulty of providing a once-for-all exact definition of one concept, whereas the latter contribution by Parkinson McCarthy and Gerring (1994) shows how difficult it can be to prevent this from happening during the process of creating a knowledge organization scheme.

If the formalist view is right that scientific concepts form a delineated domain of terms with fixed meanings, then the creation of knowledge organization schemes will imply a transfer model of mapping them towards an LIS classification framework, such as the structure of a thesaurus. But if Froehlich and Van Helvoort are right in contending that scientific concepts are 'plastic' and 'fuzzy', then how can LIS impose vocabulary control (!) on indexing terms in, e.g. thesauri? And

ISSN: 2324-9773

if concepts in a domain-specific classification scheme have already been standardized and hence, controlled, what is left to the activity of vocabulary control in LIS?

WHY BOTHER WITH DEFINITIONS? RESEARCH CAN PROGRESS WITHOUT THEM: THE CASES OF VIRUS RESEARCH AND THEORETICAL HIGH ENERGY PHYSICS

The American psycholinguist George Rey (1983, pp. 255-56) argues that the correct definition of a concept need not be known by the concept's competent users, and that “correctness” in concept definition is not equivalent to some kind of essential truth, but is merely provided by the optimal knowledge of it. This view implies i.a. that even though a correct definition of a concept may be available, scientists in a specific knowledge domain need not know or even agree on the definition of a concept. Similarly, the American philosopher of science, Dudley Shapere (1992) points out that no aspect of a scientific concept is “essential”, unchangeable over time, and that scientists “often do not fully understand the concepts or theories we are using: we must often explore our theories in order to discover what they involve” (Shapere, 1992, p. 295).

The review of virus research in the 20th century given by Van Helvoort (1994) is a very relevant case in this context in that it demonstrates, how the term 'virus' was applied, not primarily to denote a shared concept, but rather a shared, and evolving domain of research, where the term functioned as a kind of axis around which different research programmes were created and conducted:

> In the first half of the twentieth century the concept of virus was used, on the one hand, in the form of the concept of 'filterable virus', to demarcate a specific domain from other etiological agents such as bacteria and protozoa. On the other hand, it was used in the form of the concept of 'virus disease' to label a specific category of infectious diseases.
>
> Thus, the term 'virus' was used for a heterogenous group of meanings. In order to clarify what was meant by the concept 'virus', a new term for the infectious particle itself was introduced at the end of the 1950's: the virion. (Van Helvoort, 1994, p.221).

It appears that the 'plasticity' of the virus concept was constructed and deconstructed according to the evolutions and changes in knowledge in the field of immunology. Further, the concept functioned as a device for delineating virus researchers from other research communities in the field. These discoveries by Van Helvoort indicate that virus research was, and maybe still is, an evolving and not fully grown field of research. In Shapere's words “there is an 'openness' to scientific concepts, in that any component thereof may be extended in various ways, new components may be added, and others dropped” (Shapere, 1992, p. 295).

Another relevant case showing the openness or plasticity of scientific concepts is a questionnaire survey of 35 theoretical high energy physicists in the field of elementary particle physics by Hohenester et al (1988). Their study investigated the ways these scientists think about conceptualization of scientific communication and ideas by letting them elaborate on the scientific concepts 'Theory' and 'Model'. From the outset, the study emphasized that the objective was not to

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

arrive at formalized definitions, but to understand the methods and practice by which science proceeds.

The study found that rather than distinguishing between semantic aspects of 'theory' and 'model', the distinction made by some of the scientists was based on value criteria, i.e. a theory was regarded to be at a higher level than a model ("If I do it, it is a theory, if the other guy does it, it is a model"). The study concludes that the scientific methods are learned from practice, rather than from textbooks, and that "the knowledge and usage of the scientific method is more intuitive and implicit in scientists than it is verbalized, articulated, systematically described and then 'applied'". In Polanyi's terms, conceptual knowledge in an area of expertise is interwoven with the practice of the expert and is present in her work as 'tacit' knowledge (Polanyi, 1966). Further, we have earlier found (Hjørland and Albrechtsen, 1994) that textbooks, comprising e.g. definitions of domain-specific concepts and methods for scientific work, which are often applied to elicit domain-specific knowledge for expert systems in AI, are mostly poor and trivial sources for concept acquisition. In other words, formalization of scientific concepts in, e.g. textbooks, does not guarantee knowledge about the meaning of the concepts, because this meaning is constructed during the progress of scientific work. Hence, a transfer model of mapping formal definitions found in textbooks and other written 'authoritative' sources is a questionable approach to apply as a single strategy for concept acquisition in LIS knowledge organization schemes.

DO SCIENTIFIC CONCEPTS EVADE DEFINITIONS? THE CASE OF THE 'RELEVANCE' CONCEPT IN LIS

From a close look at relevance research in Library and Information Science (LIS), Froehlich (1994, p. 128) reaches a conclusion which is close to Van Helvoort's conclusion of the plasticity of the scientific concept 'virus':

> Relevance judgements do not conform to the professional, research distinction between relevance and pertinence. Given that relevance is a natural category, one acquired in experience, one should attend to its use in a natural language context, just as end-users might use it. In the history of information science, there is generated a dubious distinction between relevance (public, objective) and pertinence (private, subjective), a distinction made by academics and researchers undertaking a spirit of Cartesian modeling and natural science. The professionals and theoreticians forget that no matter what distinctions they make, users will use words as they acquire them...If one were forced to pick a word between the two to describe users' actual use of 'relevance', it would have to be 'pertinence', since as most of the articles [on relevance and pertinence, reviewed by Froehlich] have indicated, real-users make judgments based on their needs, not on their stated queries.

Froehlich finds that the difference in the meanings of 'relevance' and 'pertinence', as perceived by different researchers and IR-system users, is not to be regarded as a dichotomy, but as perspectives embedded in a much larger category of concepts (a 'graded', or fuzzy category with 'topic' as a core concept), which will have different forms of manifestation and be subject to different interpretations, according to the norms or criteria, that are shared by those who work in a particular

ISSN: 2324-9773

knowledge domain. We agree with Froehlich in his observations of the domain-specificity of 'relevance'. However, while it is true that one cannot stipulate a generic definition of this concept, we question his proposal of applying the prototype theory (Lakoff, 1987; Rosch, 1978) for LIS concepts, because this theory of cognitive categorization makes no room for the evolving construction of meaning in ordinary discourse, let alone the historical developments of knowledge in research. 'Plasticity' or 'fuzziness' of scientific concepts may indicate a rather immature field, which can certainly progress and evolve without rigid concept definitions (cf. the 'virus'-case). However, the existence or prevalence of 'plastic' or 'fuzzy' concepts in scientific discourse can also be due to different, sometimes implicit research paradigms in one discipline, as demonstrated by Froehlich (1994) for LIS. In our view, it is not impossible to expect that LIS will be able to construct a more fixed set of well-defined LIS concepts, but that presupposes a view of LIS, departing from a sociology of science perspective, i.e. that concepts are defined according to their communicative function in the LIS profession and research. In other words, we advocate that rather than mashing different views of scientific concepts through the 'blender-approach' of prototype theory, these different views should be kept as separate ingredients and be explicated in an LIS knowledge organization scheme.

THE DISCOURSE OF CONCEPT DEFINITION: THE CASE OF DSM-IV.

The American philosopher of science, Dudley Shapere (1984) claims that "as science proceeds, the connection between knowledge-claims, domain groupings, and descriptions (and often naming) tend to become tighter and tighter" (Shapere, 1984, p. 324). His exposition of the development of scientific concepts and delineation of a knowledge field in tight interconnection with development of knowledge is very relevant as an explanatory framework for the construction of scientific concepts: research fields can be more or less mature, and in 'immature' fields, scientific concepts can sometimes be ill-defined, or 'plastic'.

However, that the progress in knowledge, domain-groupings and construction of scientific concepts, which is argued by Shapere (1984, p. 324), does not always follow such ideal paths of development, and that 'plastic' concepts are not totally acceptable when a profession initiates progress in research, is demonstrated by Parkinson McCarthy and Gerring (1994) in their discourse analysis of the new version of the psychiatrical classification scheme for mental disorders, DSM-IV (Diagnostic and Statistical Manual). They claim that the creators of DSM-IV intend to force a premature birth of an authoritative category structure, independent of the development of (different) knowledge claims, and designed, not as a tool for communication between researchers and practitioners working in this knowledge domain, but as a competitive parameter, functioning as a control device for the progress in knowledge in the psychiatrical profession as well as in competing professions, such as clinical psychology, by attempting to provide these interested parties with formalized definitions of what mental disorders are. The discourse analysis of DSM-IV is conducted via i) a historical exploration of the predecessors of DSM-V; ii) an interpretation of DSM-IV's task force leaders' published writings about the revision process; and iii) a naturalistic investigation of the revision work in one of the subcommittees: The Eating Disorder Work Group.

Re i) The articulated purpose of publishing DSM-IV and its predecessors is to provide a worldwide classification system which can function as a *standard* for diagnosing and reporting mental

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

disorders. However, its immediate predecessors, DSM-III and DSM-IIIR, are found to represent a shift in psychiatry towards 'the biomedical model' of mental disorders, where these are seen as discrete phenomena, caused by biological factors in individuals. Parkinson McCarthy and Gerring (1994) trace this shift back to the German psychiatrist Emil Kraepelin who presented a textbook, intended as a cognitive authority for students and practitioners in psychiatry, comprising a classification system of 48 disorders with formal definitions of each. Until the advent of DSM-III in 1980, such approaches were unsuccessful, however, in having any impact in the United States. But immediately after DSM-III's publication, a hot debate between researchers and practitioners in psychiatry and clinical psychology evolved, and in 1992, the creation of DSM-IV began. Parkinson McCarthy & Gerring (1994) claim that this revision activity was not an incidental event, and "we believe the revision is best understood as working to control the shape of the intertextual web of the mental health field and thus influencing the agenda for research as well as relationships among contending constituencies" (Parkinson McCarthy & Gerring, 1994, p. 163). By 'intertextual web' is meant the communication between the competing authors of research papers on how to conceive mental disorders, as found in their citation patterns.

Re ii) In their published papers on DSM-IV, the task force leaders of DSM-IV are found to more or less explicitly to promote a "Progress of Science Narrative", i.e. that the revision is following an 'objective' and 'neutral' course of scientific development, based on empirical investigations of each disorder in the scheme. According to Parkinson McCarthy & Gerring (1994) this implies that DSM-IV is ultimately put forward as an agenda for basic research in the field and is placed on top of three knowledge levels in psychiatry: basic research, applied research, practice.

Re iii): The naturalistic investigation identified some conflicting views of how concept acquisition should be done for DSM-IV: One member of the subcommittee on Eating Disorders tried to promote a new category: Binge Eating Disorder ('Overeating') for DSM-IV, identified in his clinical practice. The task force leader for DSM-IV challenged him to provide empirical evidence for this new disorder, and in so doing, he actually implied the possibility that DSM-IV can precede research. In contrast, another member of the committee would rather have DSM-IV follow research and "wants to establish psychiatry as a mature scientific field beyond such controversy, beyond the place where pressure from lay groups can influence what counts as a mental disorder" (p. 176). As a compromise between the empiricist approach to category identification and an expert consensus model for acknowledging the new concept, the proposer was asked to conduct empirical research in Binge Eating and publish the results in acknowledged medical journals. The members of the subcommittee were offered a position as authorities for assessing the new category. Eventually, the proposer succeeded in getting funding for his research, published his work together with experts in the field of eating disorders (one of them on the subcommittee), and the new concept was entered as a candidate category for DSM-IV.

If this analysis by Parkinson McCarthy and Gerring can be relied upon, then their method of discourse analysis for the case of DSM-IV is of great importance to LIS for assessing the validity and quality of knowledge organization schemes created by specific professions. The discourse analysis of DSM-IV shows how standardization, and hence, formalization of scientific concepts, can be applied to transform a practice domain into a theoretical domain, which is delineated from, and even supervenes competing professions. In standardization, 'fuzzy' or 'plastic' concepts cannot be accepted. Their definitions shall be unambiguous, and there is no room for dissent. In this case,

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

consensus cannot be arrived at by adhering to right of the better argument, forwarded by the German sociologist and language philosopher Jürgen Habermas (1971), because every concept should be subject to empirical evidence. LIS should hence be careful about taking over existing professional/scientific classifications without critical investigations of the history of their social construction, and: “The criticism of a scientific categorization normally implies a critique of the science, which has developed that categorization.” (Hjørland, 1994, p. 96).

DISCOURSE ANALYSIS MADE EASY:
THE CASE OF AUTOMATED TEXT ANALYSIS FOR IMMUNOLOGY RESEARCH

Parkinson McCarthy and Gerring's analysis of the discourse in the creation of DSM-IV represents an interpretative approach to identifying conflicts in the construction and use of scientific concepts. Discourse analysis can, however, also be understood in a formalist sense, namely as an automatic text analysis of text corpora compiled from one domain. Zelig Harris (formalist language school) proposes a metalanguage for formalizing what he calls “science sublanguages” (Harris, 1988, pp. 33-56). He presents a project on 'discourse analysis' of the literature of immunology (c1935-1966), where the method used for eliciting the relevant “science sublanguage” was to compile a text corpus with linguistic data from the articles, analyze word co-occurrences and identify synonyms, resulting in a classificatory structure, equivalent to domain-specific facets: AGENTS and OBJECTS: antigen(G), antibody (A), inject (J), tissue (T), cell (C), and ACTIONS and PROCESSES between specific agents and objects, such as: appear in, produced by, secreted by, between A and C. (Harris, pp. 41-48). The purpose of creating this structure was to obtain an overall framework for assessing the field of immunology, e.g. locate differences in experimental designs and disagreements about e.g. whether lymphocytes produce antibodies or not. Harris's data-driven approach, though very interesting seen from a domain-analytic perspective, explicitly ignores what is termed as 'metascience material': “Metascience material, giving the scientist's relation to the information of the science, can be separated off” (Harris, 1988, p. 45). This reveals a mechanistic conception of scientific communication, in that it did not consider the societal conditions and constraints under which the immunology research was conducted, nor did it consider any possible evolution, change or plasticity of the terms in the field, as demonstrated by Van Helvoort for the term 'virus', a term belonging to a related research domain.

Harris's work is, however, very relevant in the context of classification practice, in particular for concept acquisition in LIS thesauri, which is often conducted using similar techniques, usually termed as “bottom-up thesaurus construction” (cf. e.g. Lancaster, 1977); such techniques have paved the way for semi-automatic thesaurus construction methods, where terminology compilation and facet analysis follow similar principles. In this context it should be mentioned that Harris's work (e.g. Harris, 1968 and 1988) is viewed (for instance by Jakobson, 1973) as classic in the field of mathematical linguistics, where set theory, Boolean algebra, statistic calculus of probability and information theory from mathematics have been applied to study the structure of language. It is thought-provoking to consider how applications of mathematical theories in linguistics have been transferred to LIS practice without discussions of their origin in mathematics, nor their possible shortcomings for language processing, let alone discussions of their applicability for LIS requirements.

ISSN: 2324-9773

DISCUSSION

It appears from the above discussions in cognitive psychology and linguistics, sociolinguistics and philosophy of science of scientific concept formation that scientific concepts and terms play different roles depending on the societal and developmental conditions in a particular field. What lessons can be learned from this?

i) A field may be immature, as is perhaps the case for the virus concept, resulting in 'plastic' concepts, because even though a field sets out to study a particular phenomenon, it only knows something, not all about this phenomenon — that is the very aim of the research — but the field has to give its 'child' a name, and hence this 'name', or rather term will have different meanings over time. ii) A field may be mature, perhaps, as in the case of high energy physics, but engaging in concept definitions is secondary to using the concepts as tools in scientific activity and communication, because the field progresses so fast that any definition will lag behind; iii) 'Fuzziness' in scientific concepts can also be due to competing, sometimes implicit, research paradigms in one field, such as LIS, where extremities include rationalistic definition endeavours/ empiricist data compilation versus studies of individual users' cognitive categorizations. In this case, one concept will have different meanings; iv) Classification schemes, such as DSM-III and DSM-IV, can be promoted as an authoritative source of how a field is organized, but may disregard other worldviews of the same domain; hence, one concept may be presented as having a fixed, standard meaning, but may in fact have different meanings, which should be found elsewhere; v) From the formalistic study by Harris in immunology we find that even though terms and structures may be acquired automatically, the scientific concepts lose their metascientific contextualizations, providing the very framework for interpreting the data.

It is hence not an easy task to stipulate a ready-to-hand technique of identifying and analyzing scientific concepts for concept acquisition in LIS knowledge organization schemes. For each domain investigated, the methods for doing this will vary. But from the cases presented above we find that interpretative approaches, based on critical assessments of existing classification schemes and discussions of concepts in review literature, are very good sources for gaining knowledge about concepts, worldviews and the progress of knowledge in a domain. Scientific concepts may perhaps to some extent be identified automatically, but the context of the compiled data may be lost. *What can and should be investigated is the communicative function, the variety and history of scientific concepts.*

WHEN AND WHY DID ORDINARY LANGUAGE BECOME EXTRAORDINARY?

The reasons for variations in the meanings of scientific concepts cannot, however, solely be explained by introducing a philosophy or sociology of science perspective, as we have implied above. A deeper understanding of the problems at stake entails a discussion of different conceptions of language, i.e. formalist versus more hermeneutical conceptions of language and their implications for whether to distrust the scientific organization of knowledge as artificial in favour of promoting more informal, or "natural" approaches to the organization of knowledge.

In his article "How Ordinary is Ordinary Language?" (1973), Stanley E. Fish questions the method of text analysis, based on linguistics, prevailing in the 1970's in literature studies, where some literature critics propose a distinction between literary and ordinary language, in attempts at setting

ISSN: 2324-9773

up boundaries between the profession of literary critics (whose object is literary language) and linguists (whose object is ordinary language). One critic even concludes that "The fact of the matter is that criticism...is an autonomous activity". In Fish's view, this attempt at establishing literary critics as an autonomous institution i) disregards the interplay between linguistics and literature ii) accepts "the positivist assumption that ordinary language is available to a purely formalist description", hence excluding *the subjective aspects of literature, e.g. values and intentions* iii) creates a deviation theory for both, which "trivializes the norm and therefore trivializes everything else" (Fish, 1973, p. 44).

Fish concludes that "*There is no such thing as ordinary language*, at least not in the naive sense often intended by that the term: an abstract formal system, which, in John Searle's words, is only used incidentally for purposes of human communication" (Fish, 1973, p.49). Fish's point that 'ordinary language' has been deliberately segregated from professional discourse as an object for formal language analysis, a segregation which is rooted in a positivist, formalist conception of language, indicates that the dualism between the two language domains is questionable, because it has been imposed rather than discovered. In her critique of contemporary terminology work in Denmark, distinguishing between Languages for Special Purposes (LSP) and Languages for General Purposes (GLP), the Danish linguist Carol Henriksen thus concludes that "Every time you use language, it is for a special purpose." (Henriksen, 1990, p. 28).

A number of studies of academic theory versus commonplace understanding in various knowledge domains address the relationship of understanding the lay level with that of the scientist, where some authors taking a social constructionist perspective of language argue that scientists often draw on roughly the same domain of everyday understandings as other members of the culture, thereby eradicating the traditional boundaries that have kept the two distinct (Semin and Gergen, 1992). The dualism between ordinary language and scientific language is hence an artificial distinction, imposed by a formalist view of language as an object related to different and segregated "speech communities".

In agreement with Fish (1973), Henriksen (1990) and Semin & Gergen (1992), we find that there is a dialectical interplay between everyday discourse and scientific discourse, in that in each domain, language is used for a specific purpose. The interweb between the two discourses is elegantly explained by the Norwegian linguist Ragnar Rommetveit (1992), who advocates a "dialogical perspective":

> Contextbound understanding of our "Lebenswelt" from within as mediated by ordinary language is transcended by fixation of perspective within separate domains of practical-technological, professional and scientific expertise *and* by exploitation of technology that extends and transcends our human sensory-motor equipment. Fixation of perspective is thus essential in the ramification of fragments of ordinary language into highly specialized technological, professional and scientific terminologies, i.e. in the historical development of monological discourse within different enclaves of expert knowledge out of initially holistic, perspective-relative, and socially negotiable human cognition. Over time, such terminologies are to some extent assimilated into the everyday language of enlightened lay people and reflected in a novel, collectively endorsed standards of correctness for use and

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research. 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

> novel, shared realities. Ordinary language as recursively affected by its transcendence under conditions of historical change thus entails significant transformations of our understanding of the world and ourselves." (Rommetveit, 1992, pp. 23-24).

CATEGORIZATION IN EVERYDAY DISCOURSE *VERSUS* SCIENTIFIC CLASSIFICATIONS.

In a recent contribution to a discussion of classification as a communicative constraint, Jacob (1994) has argued, that scientific classification is too restrictive, too constrained because of the special perspectives in the sciences, and that less domain-specific, less rigid categorizations are needed. Jacob's view of the restrictiveness of scientific classifications may imply a criticism of tendencies in these sciences:

> The need to ensure that disciplinary knowledge is consistent across individuals favor of the stability of reference provided by well-defined, discipline-based classes and forces the surrender of that very flexibility and plasticity that characterizes cognitive categories. As a result, experientially-based categories lose their inherent ability to accommodate new or individualized experiences and are transformed into rigidly bounded and concretized domain-specific classes through a process of formal analytic definition." (Jacob, 1994, p. 102).

We have earlier argued that scientific concepts are not always formalized or fixed, and that they can be applied as flexible tools to facilitate communication and hence the progress in science. At the same time, such plasticity in scientific concepts poses a challenge to vocabulary control in LIS: in Jacob's terms, a traditional LIS classification scheme is artificial in that "it is a tool or artifact created to express purpose of establishing order". This is put in opposition to the process of communication between individuals: "This process of becoming demands that the meaning of the word is not concretized but remains flexible, plastic, and responsive, capable of incorporating, or being incorporated within, the as yet unspoken response." We agree that LIS classification schemes are often rigid in their structures and concepts definitions, but these schemes have not been built as stand-alone artifacts, but as tools for retrieval. However, we find that the problem facing LIS may be its tendency to acquire scientific concepts, that have been formally defined, hence implicitly overtaking a scientistic worldview. In addition, one may also question the rigidity of the hierarchies or network structures of delineated concepts which characterizes the traditional approaches to classification and thesaurus construction. In short, the structuralist paradigm imposed on LIS, founded on a Saussurean view of language as a sign system and transferred (implicitly) to, in particular thesauri: "A linguistic sign, [Saussure] suggested, may therefore adequately be defined only in terms of the *relations* which it contracts with other signs: linguistic units possess a purely relational identity" (Sinha, 1988, p. 19). In other words, we propose that LIS should pay more attention to the middle level between traditional classification schemes and the communication between individuals: the dynamic construction of scientific concepts and the challenges that this poses for defining more flexible structures and concept definitions in LIS representations of knowledge. This imposes a requirement for LIS to question its all-embracing practice of structural analysis, in casu facet analysis of concepts, and pay more attention than hitherto to concept analysis

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

using an interpretative approach of identifying worldviews, concept development and linguistic variations.

LANGUAGE, CATEGORIZATION AND CLASSIFICATION

The famous American linguist Leonard Bloomfield wrote in his main work "Language" from 1933, that the correct meaning of a word is the scientific analysis of that word. He used the example "salt" and argued, that the chemical analysis of salt as "NaCl" determined the exact meaning of this word. Bloomfield represented a view of language of an extreme behavioristic, positivistic and nominalistic nature. Nominalism is the view, that concepts are only labels put on individual perceptions of units of knowledge. Nominalism makes no room for an active role for language in the perception of reality. This view is now widely regarded as obsolete. Today, influences from e.g. hermeneutics have shown, that language does play an active role in cognition. Among the arguments raised against Bloomfield's nominalism, is the following by the Danish linguist Lars Henriksen (1994): If Bloomfield was right, then it would be impossible to speak about tulips without asking a botanist, or to say that it is raining without being a meteorologist. A scientific explanation is not the same as the meaning of words, for the simple reason, that words do not indicate reality - but the experienced reality. The words express our experience and understanding of reality, not its objective and true nature. In the latter case, words like "mermaid" and "unicorn" would be meaningless, because their referents are not existing, concrete phenomena.

The rejection of nominalistic positions such as Bloomfield's has promoted more subjective and relativistic theories, including theories from hermeneutical traditions, focussing on the communicative function of language (e.g. Halliday, 1977, Jakobson, 1973), the culture-specific aspects of language of language structure and use (e.g. Berlin & Kay, 1969, Whorf, 1965) and the experiential aspects of everyday categorization (Lakoff, 1987). Where nominalism made no place for language in the perception of reality, some of the alternative theories have even defined language something all-embracing or comprehensive, e.g.: "Nothing exists except through language", Winograd & Flores argue (1987, e.g. p.73).

We have argued that there is a relationship as well as a distinction between ordinary and scientific language: in both domains, meaning is constructed and constantly evolving in oral and written communication. However, ordinary and scientific discourse are constrained by very different goals. The aim of research is to make progress in general knowledge where the goals and intentions of ordinary discourse are specific to the concrete activity in which its participants are involved. The creation and meaning of concepts in everyday discourse is hence of a situation-specific kind, whereas the formation and organization of scientific knowledge is of a general kind. In indicating that scientific principles are the most general kind of knowledge organization, have we then returned to the behaviorism of Bloomfield? No, we have not. But we have also tried to avoid the relativity and subjectivity of other views, including many hermeneutic approaches. Theories about mind and language based on socio-cognitive and domain-specific principles have an increasing influence now in the 1990's after the domination of very rationalistic, AI-like theories in the 1980's. The Norwegian psycholinguist Ragnar Rommetveit is one important researcher, who is following socio-cognitive and dialogical principles. In his view, the important difference between ordinary and scientific language is thus the fixedness of perspective in scientific language, and the more

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

flexible, unbound meanings in ordinary language. Both kinds of languages are effective in their respective domains (Rommetveit, 1992).

IMPLICATIONS FOR CLASSIFICATION RESEARCH OR: HOW CONTROLLED ARE CONTROLLED VOCABULARIES?

The main thesis of this contribution was that the duality between scientific and ordinary language is questionable, and we found that this duality was rooted in a certain epistemological position, namely a formalist view of language, which disregards the interplay between scientific and ordinary cognition, and promotes a very rigoristic (linguistic) division of labor. We have presented the communicative aspects of scientific concept formation and found that in order to acquire such concepts for LIS classification schemes, the formalist approaches are inadequate, and should be replaced by more interpretative approaches. We have tried to clarify how scientific knowledge is of a general kind, created in the activity of research, and that there is a great variety in the meaning and use of the concepts, which is not to be found via automatic concept acquisition.

Traditionally, the activities of constructing LIS knowledge organization schemes follow a model of concept identification, concept analysis (and facet analysis), terminological control and semantic control, often resulting in a kind of 'crucifixion' approach, which constrains further developments in knowledge about alternative worldviews or structures in the activity of the thesaurus designer. Concept identification and concept analysis are fundamental to building a thesaurus, which is not "just another semiotics system" (in the terminology of Frohmann, 1994), and we shall conclude by shedding light on these activities and provide our recommendations for concept acquisition and analysis in thesauri, following an interpretative approach rather than a formalist approach.

Concept identification for LIS thesauri is usually recommended to follow a topdown approach, a bottom-up approach, or a combination of both (Lancaster, 1977). Concept identification using a topdown approach investigates existing classification schemes, thesauri and authoritative dictionaries and textbooks of the covered knowledge domain. This presupposes an expert consensus model as a point of departure. However, concepts in existing classification schemes, as demonstrated by Parkinson McCarthy and Gerring (1994) for DSM-IV in psychiatry, have not always been created using a consensus model even though such schemes may have been promoted and applied as authoritative sources. Further, there may be conflicting views of scientific concepts, and their meaning will evolve over time. Concept acquisition using a bottom-up approach is based empiricist techniques, such as compilation of terms from samples of representative texts, produced in a knowledge domain. This approach is questionable when applied in splendid isolation from discourse analysis, i.e. analysis of why and when who communicates what to whom. It is however thought-provoking that this approach, which is based on the assumption of some kind of empirical evidence, termed as "literary warrant", has hardly ever been questioned by LIS.

The topdown approach, however, has been subject to critique. Lancaster, who advocates that concept acquisition should be based on "warrant" from the literature (literary warrant) or from the information needs of the users (user warrant), warns against the topdown approach to concept acquisition:

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

> The danger of this approach, which is exemplified by the work of the Engineers Joint Council and by Project LEX is that the terms may lack true literary warrant, or more particularly, that the vocabulary will be developed in an unbalanced way, not providing variations in detail that truly reflect the varying needs of a community of users ("user warrant)". (Lancaster, 1977, p. 14)

How to arrive at "user warrant", reflecting 'the varying needs of a community of users', remains as a difficult problem facing LIS. First, it would entail the presence of knowledge of who the users of IR-systems are: disciplinary, interdisciplinary, laymen etc. Second, it would mean that we know how they would like to communicate with IR-systems. Third, we would have to know something about the 'variability' of their needs. Such a project seems rather comprehensive compared to analyzing the variety of concepts in the written, e.g. scientific, literature, and expose this information to the users in thesauri, providing them with a map which can be browsed as a ready-to-hand knowledge source. One example of such a map is the Unified Medical Language System (UMLS) (Schuyler et al, 1993), which exposes different views of biomedical concepts with scope notes of their origin in different sources (classification schemes and thesauri). The latter approach does not, however, indicate why and how these varieties have evolved, and further, it is a metathesaurus, not a thesaurus which is applied by indexers as well as by users.

Concept analysis usually follows an approach of identifying generic categories of concepts, such as actions or events, which are often applied as a basis for facet analysis and logical division of terms. This approach is questioned by i.a. Frohmann:

> The concept organization may be so different in different subject areas to justify treating, for the purposes of classification, the relevant generic terms as standing for different kinds of entity. It follows that the various hierarchies must be organized by reference to the consensus or the problems within the relevant subject field...when terms are used in different practices, the semantic relations between them will differ accordingly" (Frohmann, 1983, p. 17)

When a thesaurus designer bases her work on literary or user warrant for concept acquisition, a major problem remains: how to perform semantic control, i.e. devise hierarchies and related term relationships. Terms compiled from text corpora form an atomized amount of entities whose semantic, contextual, communicative aspects have been lost — even though it is to sometimes possible to identify semantic relationships via analysis of term co-occurrences. This means that the designer is left with her intuitions or formal logic for creating such structures. We agree with Frohmann's recommendations of adhering to consensus or problems within the relevant knowledge domain. Otherwise the activity of semantic control will be conducted without guidance from the lessons learned from concept acquisition.

We find that LIS is focussing very much on the formal, technical aspects of thesaurus construction and needs to gain more guidance for the interpretative aspect of this activity. The formalist approach is reflected in the practice of adhering to standard and generic recommendations for this task (ISO, 1985), as pointed out by Svenonius (1986) and Krook & Lancaster (1993). The focus is how to normalize and disambiguate the richness of natural language offered for free-text searching in databases to fit the slimmer model of a thesaurus. Or, in the terminology of Frohmann (1994) to

ISSN: 2324-9773

build a classification system, in casu DDC, as a stand-alone, purely semiotic system of call numbers, which does not have any referent beyond itself. So we have two extreme models: i) thesauri based on literary warrant referring to terms acquired at a certain point of time from a certain text corpus ii) classification schemes built as stand-alone semiotic systems to which any given document must submit. LIS (in particular Lancaster, 1977) offers one middle position between these extremes in what he calls a "deductive method" for concept acquisition: to index a number of representative documents and use the indexing terms as candidate thesaurus terms. He thus implicitly introduces the importance of subjects in documents and in so doing, leaves the responsibility for concept analysis to the indexers. What is lacking is a guidance for finding this middle level of subjects in the gap between literary warrant and the construction of scientific concepts. This guidance can be provided by adhering to an interpretative approach to constructing knowledge organization scheme.

We propose to rephrase the traditional "concept identification, concept analysis (and facet analysis), terminological control, semantic control-model" as:

- Concept investigation and interpretation
- Choice of structures
- Concept articulation

Concept investigation and interpretation cover critical analyses of concept formations, variations and developments rather than compilation of terms from documents (empiricist model) or promotions of stand-alone semiotic systems (formalist control model). This imposes a more epistemological and pragmatic level to this activity, which identifies conflicting views, arguments etc., where facet analysis decomposes concepts into primitive constituents which — in isolation — have no meaning.

Structures should be less rigid than traditionally, at least for thesauri. In some cases, a structure of clustered concepts will be more communicative to the searcher than conventional hierachical/ network structures, in particular because a rigid structure will appear rather artificial in a field with many new and evolving concepts.

Concept articulation comprises an explanation and exposition of the knowledge about the concepts, found during concept investigation and interpretation. This is suggested as a broadening of the useful scope note facility often found in thesauri. Such extended scope notes can be of different kinds: Explaining how a concept has evolved over time, exposing different conceptions of the same concept, etc., hence providing a more pluralistic map of concepts for browsing than has hitherto been provided. UMLS's project of mapping medical concepts from different Knowledge Organization sources is one step in this direction (Schuyler et al, 1993). However, for metathesauri, each K.O. source is in itself founded on different worldviews and has been produced under certain constraints. For instance, the predecessor of DSM-IV, which has been analyses to support a specific exclusive worldview, is part of the knowledge sources in UMLS, and such information should be explicated to the searcher. Our recommendations for how to develop LIS knowledge organization schemes, taking an interpretative rather than a formalist approach, should be seen as an argument for more emphasis on the aspect of knowing rather than constructing. We believe that this knowing on the part of the developer can function to provide more useful, flexible and varied knowledge organization schemes which can support the choice of search strategies and developments in knowledge on the part of the searcher.

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

REFERENCES

Berlin, B. and Kay, P. (1969). *Basic Color Terms*. Berkeley and Los Angeles (CA): University of California Press.

Brier, Søren (1992). A Philosophy of science perspective on the idea of a unifying information science. (in: Pertti Vakkari & Blaise Cronin (eds.): *Conceptions of Library and Information Science. Historical, empirical and theoretical perspectives*. Tampere, COLIS Conference, 1991. London: Taylor Graham, pp. 97-108.

Christensen, Marianne L. (1994). Hermeneutik - fortolkning og forståelse. *Biblioteksarbejde* 15 (41), pp. 25-40 (in Danish: Hermeneutics: Interpretation and Understanding)

Fish, Stanley E. (1973). How Ordinary is Ordinary Language? *New Literary History* 5(1), pp.41-54

Froehlich, Thomas J.(1994). Relevance Reconsidered - Towards an Agenda for the 21st Century: Introduction to Special Topic Issue on Relevance Research. *Journal of the American Society for Information Science (JASIS)* 45(3), pp. 124-134.

Frohmann, Bernard P.(1983). An Investigation of the Semantic Basis of Some Theoretical Principles of Classification Proposed by Austin and the CRG. *Cataloging & Classification Quarterly* 4(1), pp. 11-27.

Frohmann, Bernard (1994). The Social Construction of Knowledge Organization: The Case of Melvil Dewey. (in: Hanne Albrechtsen and Susanne Ørnager (eds): *Knowledge Organization and Quality Management. Proceedings of the 3rd International ISKO Conference, 21.-24 June 1994*. Frankfurt: Index Verlag, 1994, pp. 109-117).

Gregersen, Frans (1990). Historicitet og situation. *Psyke og Logos* 11, pp. 121-137. (In Danish: Historicity and Situation)

Habermas, Jürgen (1971). Vorbereitende Bemerkungen zu einer Theorie der kommunikativen Kompetenz (in: Habermas & Luhmann, eds.: *Theorie der Gesellschafft oder Socialtechnologie*. Frankfurt am Main: Suhrkamp, pp. 101-142) (In German: Prolegomena to a Theory of Communicative Competence)

Halliday, M.A.K. (1977). *Explorations in the Functions of Language*. Amsterdam: Elsevier. 144p.

Harris, Zelig (1988). *Language and Information*. New York: Columbia University Press. ix,120 p.

Harris, Zelig (1968). *Mathematical Structures of Language*. New York: Krieger. 240 p.

Henriksen, Carol (1990). *Two Papers on "Fag(sprog)lig kommunikation" (Language for Specific Purposes)*. Roskilde: Roskilde Universitetscenter. 49 p. (ED324970)

Henriksen, Lars: LOKUS: Der hvorfra et ord ser sit emne. *Sprint. Sproginstitutternes Tidsskrift. Handelshøjskolen i København*, 1994,(1), pp. 33-44. (In Danish: From where a word views its subject).

Hjørland, Birger (1994). Nine Principles of Knowledge Organization (in: Hanne Albrechtsen & Susanne Ornager (eds): *Knowledge Organization and Quality Management. Proceedings of the 3rd International ISKO Conference, 21.-24 June 1994*. Frankfurt: Index Verlag, 1994, pp. 91-100.

Hjørland, Birger & Hanne Albrechtsen (1994). *Toward A New Horizon in Information Science: Domain Analysis*. Manuscript delivered to a major IS journal.

Hjørland, Birger (1993): *Emnerepraesentation og informationssøgning. Bidrag til en teori på kundskabsteoretisk grundlag*. Göteborg: Valfrid. 259 p.(English summary: Subject Representation and Information Seeking. Contributions to a Theory based on the Theory of Knowledge. (Pp 216-223).

Hjørland, Birger (1992): The Concept of "Subject" in Information Science. *Journal of Documentation*, 48(2), page 172-200.

ISSN: 2324-9773

Albrechtson, H. (1994). Understandings of Language and Cognition: Implications for Classification Research . 5th ASIS SIG/CR Classification Research Workshop, 1-16. doi:10.7152/acro.v5i1.13774

Hohenester, A., L. Mathelitsch and M.J. Moravcsik (1988). The Usage of 'Theory' and 'Model' in Scientific Conceptualization. *Scientometrics* 14, pp. 411-420.

ISO (1984). *Guidelines for the Establishment and Development of Monolingual Thesauri.* International Organization for Standardization (ISO 2788).

Jacob, Elin K. (1994). Classification and Crossdisciplinary Communication: Breaching the Boundaries Imposed by Classificatory Structure. (in: Hanne Albrechtsen & Susanne lrnager (eds): *Knowledge Organization and Quality Management. Proceedings of the 3rd International ISKO Conference, 21.-24 June 1994.* Frankfurt: Index Verlag, 1994, pp. 101-108.)

Jakobsen, Roman (1973). *Main Trends in the Science of Language.* London: George Allen & Unwin. 75 p.

Krook, David A. & F.W. Lancaster (1993). The Evolution of Guidelines for Thesaurus Construction. *Libri* 43(4), pp. 326-342

Lakoff, George. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press. xviii, 614 p.

Lancaster, F. W. (1977).Vocabulary Control in Information Retrieval Systems. (in: *Advances In Librarianship* / edited by Melvin J. Voigt and Michael H. Harris. - New York: Academic Press, pp. 2-40).

Parkinson McCarthy, Lucille & Joan Page Gerring (1994). Revising Psychiatry's Charter Document DSM-IV. *Written Communication* 11(2), pp. 147-192.

Polanyi, Michael (1966). *The Tacit Dimension.* New York: Doubleday & Company.

Rey, Georges (1983). Concepts and Stereotypes. *Cognition* 15, pp. 237-362.

Rommetveit, Ragnar (1992). Outline of a dialogically based social-cognitive approach to human cognition and communication. (in: Astri Heen Wold (ed.): *The Dialogical Alternative. Towards a Theory of Language and Mind.* Oslo: Scandinavian University Press, 1992, pp. 19-44.

Rosch, Elinor (1978). *Cognition and Categorization.* Hillsdale (NJ): Erlbaum.

Schuyler, Peri L., William T. Hole, Mark S. Tuttle, David D. Sherertz (1993). The UMLS Metathesaurus: Representing Different Views of Biomedical Concepts. *Bulletin of the Medical Library Association* 81(2), pp. 217-222

Sermin, Gün R. & Kenneth J. Gergen (eds.) (1990). *Everyday Understanding: Social and Scientific Implications.* Newbury Park (CA): Sage. 248 p.

Shapere, Dudley (1992). Talking and Thinking about Nature: Roots, Evolution, Future Prospects. *Dialectica* 46 (3-4), pp. 281-296.

Shapere, Dudley (1984). *Reason and the Search for Knowledge. Investigations in the Philosophy of Science.* Dordrecht: D.Reidel Publ. Comp.(Boston Studies in the Philosophy of Science; v. 78)

Sinha, Chris (1988). *Language and Representation: a Socio-Naturalistic Approach to Human Development.* New York: Harvester/Wheatsheaf. xix,235 p.

Svenonius, Elaine (1986). Unanswered Questions in the Design of Controlled Vocabularies. *Journal of the American Society for Information Science* 37(5), pp. 331-340.

Van Helvoort, Ton (1994). History of Virus Research in the Twentieth Century: The Problem of Conceptual Continuity. *History of Science* 32(2), pp. 185-235.

Whorf, B.L. (1956). *Language, Thought and Reality.* Boston: Massachusetts Institute of Technology Press.

ISSN: 2324-9773