The foundation of LIS in information science and semiotics¹

by Søren Brier

Introducing the conflict between informational and semiotic paradigms

wo major strategies for gaining a systematic understanding of the "laws" of information, cognition and signification and communication are the informational and the semiotic. They are both transdisciplinary and universal in scope, but they study the basic ideas of information, cognition and communication from disparate angles. Nöth writes about the relationship between these paradigms:

Information in its everyday sense is a qualitative concept associated with meaning and news. However, in the theory of information, it is a technical term, which describes only quantifiable aspects of messages. Information theory and semiotics have goals of similar analytic universality: Both study messages of any kind, but because of its strictly quantitative approach, information theory is much more restrictive in its scope.

(Nöth 1995: 34)

This article states the conflict between informational and semiotic approaches to cognition and communication and points to crucial differences in the metaphysical framework behind the pan-informational and the pan-semiotic paradigms as one of the obstacles for making a transdisciplinary framework integrating them in search of a theoretical framework that can encompass truth and meaning, science and humanities. We will then take this problem deep into some basic practical problems of subject searching in library and information science

¹ The present article is based on material from the following articles: Brier, S. (1992): "A philosophy of science perspective - on the idea of a unifying information science". In P. Vakkari and B. Cronin (eds.) (1992): Conceptions of Library and Information Science: Historical, empirical and theoretical perspectives, Taylor Graham, London, pp.97-108. Brier, S. (1996): "Cybersemiotics: A new interdisciplinary development applied to the problems of knowledge organization and document retrieval in information sci-ence", Journal of Documentation, Vol. 52, no. 3, September 1996, pp. 296-344. Brier, S. (1996):"The Usefulness of Cybersemiotics in dealing with Problems of Knowledge Organization and Document Mediating Systems", Cybernetics: Quarterly Review of the International Association for Cybernetics, Vol. XXXIX, no. 4. pp. 273-299. Brier, S. (2000): "Trans-Scientific Frameworks of Knowing: Complementarity Views of the Different Types of Human Knowledge", Yearbook Edition of Systems Research & Behavioral Science. System Research, 17, pp. 433-458.Brier, S. (2000b). Brier, S. (2001): "Cybersemiotics: A Reconceptualization of the Foundation for Information Science ", Systems Research and Behavioral Science, Yearbook. Systems Research 18, pp.421-427. Thellefsen, T.L., Brier, S. and Thellefsen, M.L. (2003): "Problems concerning the process of subject analysis and the practice of indexing: A Peircian semiotic and semantic approach toward user oriented needs in document searching," Semiotica, 144-1/4 (2003), pp.177-218. Brier, S.: (2004): "Cybersemiotics and the Problem of the Information-Processing Paradigm as a Candidate for a Unified Science of Information Behind Library and Information Science", pp.629-657 in Library Trends, Vol. 52, No. 3, Winter 2004.

(LIS) to show the practical limits and problems of the universal theory of objective information as the foundation of cognition and communication science.

This universal theory is often called the "information-processing paradigm". It is built on an objective information concept combined with a general idea of computation that is usually algorithmic. The mechanistic and rationalistic information-processing paradigm prevailing in cognitive science is the predominate approach in this trans-disciplinary area, which is dominated by computer science and informatics. In the analysis below, I demonstrate that the logical and mechanistic approach alone cannot offer an understanding of human signification and its basis in biological, psychological and social relationships. I then discuss the ontological and epistemological problems of the idea of "information science" by discussing information concepts and paradigms based upon other basic epistemological and ontological theories.

In discussing the possibility of a universal information science (which must include a universal science of communication and cognition) it is important to analyze the nature of subject areas that a universal information science has to combine, such as physics, biology, social science, humanities, library and information science, computer science, cybernetics, communication and linguistics. The strategies for developing an information science is to extract the areas of information, knowledge, perception and intelligence from the old philosophical tradition and its pondering about phenomenology, qualia, consciousness, meaning and signification, epistemology and ontology, and instead develop an efficient objective science called cognitive science. Such a move attempts to release us from more than two thousand years of philosophical discussions on cognition, signification and meaning, by turning the subject into an empirical science.

Many information "scientists" would claim that it is exactly this restriction that makes it possible to construct a universal theory of information and cognition. Thus the qualitative phenomenological and pragmatic approach of semiotics seems to make it unsuitable for the sciences, which are presently grounded in either mechanistic atomistic determinism or in some type of Gibbs probabilistic complexity theory (Hayles 1999 p. 88-90 analysis of Wiener's theoretical foundation and Prigogine and Stengers 1984) also influenced by the uncertainty principle in quantum mechanics.

I want to consider these differences as general philosophical and methodological problems for the study of information, cognition, signification and communication as a transdisciplinary field. The problem is basic to the entire field of information, cognitive science, signification and meaning. A basic inquiry is whether the functionalistic and cybernetic research program of information and cognitive sciences must be seen as complementary to a phenomenological-semiotic line of theorizing on signification and meaning that ignores ontological questions outside culture, or if they might be united into one paradigmatic framework by carrying through a revision of the ontological and epistemological foundation of both classical and modern science as Peirce attempts.

Essentially the mathematical theory of information defines information as merely the statistical property of a message, irrespective of its meaning. It is seen as a selection among

signals. In information theory, a signal has information when it excludes the occurrence of other signals that could have occurred instead. The quantification of information depends on the number of excluded alternatives and the probability with which a signal can be expected to occur. The informational value of a signal is calculated as the probability of occurrence in a message. What counts is the statistical rarity of signs, or rather, codes. Shannon's information theory, when used in a broader scientific sense, presumes that signals are meaningful codes established in a system of signs, such as the Morse code for the alphabet. In this situation one can relate this information concept to the quantitative side of meaningful communication without addressing the presupposed meaning that makes the calculation worth doing. But Norbert Wiener, the father of cybernetics, made a more general theory of information, saying that 'information is information, neither energy nor matter', but something real in nature being everywhere (pan-informational paradigm), and as Schrödinger (1967) showed in his book *What is life?*, is being especially crucial for living systems.

The modern versions of the pan-informational paradigm often combine functionalism with the non-equilibrium thermodynamics, non-linear systems dynamics, and deterministic chaos theory and fractal mathematics as descriptive tools. But again there are seldom systematic reflections on how they differ from a mechanistic view or on the nature of a concept of meaning and how signification arises in minds. This is a general philosophical problem in the area of "psychology" and "cognitive science". At least two of the methodologies in the area of human behavior, thinking and communication, presume that humans are meaning-producing systems. These are the phenomenological and hermeneutical qualitative approaches.

The large differences between the scientific approaches on the one side and the phenomenological-hermeneutical approaches on the other still fuel the debate as to whether psychology can ever establish itself as one science, though cognitive science and the information-processing paradigm are themselves such attempts that ignore the problems of meaning that phenomenology, hermeneutics and semiotics address.

Peircean based pan semiotics

Within the last twenty years a semiotic and communicational paradigm, largely based on Peirce's "(...) semiotics, that is, the doctrine of the essential nature and fundamental varieties of possible semiosis." (Peirce, CP: 5.448) has developed. Semiotics develops a general theory of all possible kinds of signs, their modes of signification and information, and whole behavior and properties. It studies the existence of meaningful communication in living and social systems and looks to cultural historical dynamics and evolutionary ecology for explanations of the dynamics of signification and communication. Peirce founded semiotics as a logic and scientific study of dynamic sign action in humans, their language, science and religion and other cultural products as well as sign in non-human nature. In the form of biosemiotics, this view is now penetrating biology as an alternative to both mechanistic and purely systems' dynamical explanations. Work has been undertaken in genetics, molecular biology and biochemistry (Hoffmeyer and Emmeche, 1991, Barbieri 2001), organic chemistry.

A pan-semiotic philosophy can be constructed from a few central quotations from Peirce. The first pertains to the ontological question of the basic elements of reality:

The entire universe is perfused with signs, if it is not composed exclusively of signs.

(Peirce, CP: 5.448, fn.)

In other words, in thinking we never have access to the thing in itself, but only as it appear to us through signs. Since we are living in a "semiosphere" (Hoffmeyer 1997) in our individual and collective "signification spheres", we never get "behind" the signs to "reality." So why not admit that signs are the only reality we will ever know? Even humans are only signs.

For, as the fact that every thought is a sign, taken in conjunction with the fact that life is a train of thought, proves that man is a sign; so, that every thought is an external sign, proves that man is an external sign. That is to say, the man and the external sign are identical, in the same sense in which the words homo and man are identical. Thus my language is the sum total of myself; for the man is the thought.

(Peirce, CP: 5.318)

In this view, semiotics becomes the fundamental doctrine and philosophy to grasp knowing and reality. Still, a prerequisite for signs to work as tools for cognition is a basic precoupling between the organism and the environment. One has to know where to look and what to look for in order to obtain further information from a sign. Peirce writes:

The sign can only represent the object and tell about it. It cannot furnish acquaintance with a recognition of that object. (...) It presupposes an acquaintance in order to convey some further information concerning it.

(Peirce, CP: 2.231)

The problem is whether this acquaintance presupposes certain pre-semiotic experiences as does much of hermeneutic philosophy. Regardless, in semiotics, meaning and signification do not have much to do with informational bits. The phenomenological theory established in Peirce's semiotics underlines the fact that qualia, interpretation and meaning are at least as important as the quantitative selection and measuring of bits.

In Peirce's triadic philosophy, feelings, qualia, habit formation, and signification are basic ontological constituents of reality. This suggests that the semiotic paradigm should be able to penetrate beyond chemistry and physics to "the bottom of nature". This seems to clash with basic beliefs in sizable parts of information science that seems to want to construct meaning as a bottom up procedure from a thermodynamically based information science.

We seem to have two completely distinct points of departure for these theories that both aim to be universal. The difference between the two paradigms is fundamental. The information paradigm is based on an objective, quantitative information concept working with algorithmic models of perception, cognition and communication. Semiotics, based in human language's meaningful communication, is phenomenological and dependent upon a theory of meaning.

One way of viewing the problem is to see the pan-informational paradigm as a "bottom-up" explanation and the pan-semiotic paradigm as a "top-down" explanation. One could further combine this with an epistemological viewpoint that suggests that no final scientific explanations can be given to anything in this world, including the behavior of organisms. All we have are complementary explanations that work well in different situations. We can never attain a full view.

According to this, it might be impossible to unite the two paradigms by manipulating basic definitions into unifying compromises. Instead we should continue to develop each paradigm to its fullest and then combine them as complementary views of a subject matter we can never fully grasp in any kind of unified scientific systems of concepts and laws (see Figure 1).

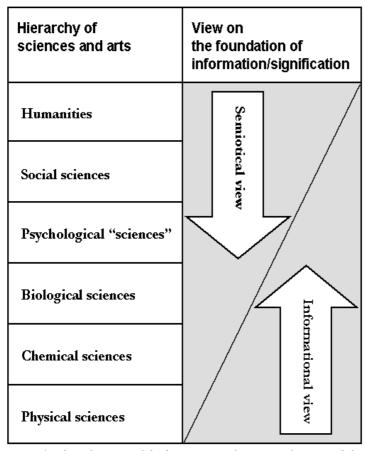


Figure 1: The relevance of the bottom-up informational view and the top-down semiotic view in the area of the foundation of information science. On the left side is a hierarchy of sciences and their objects, from physics to humanities and vice versa. On the right is an illustration of the two most common "scientific" schemes for understanding and predicting communicative and organizational behavior: 1. the semiotic top-down paradigm of signification, cognition and communicative, and 2. the informational bottom-up functionalistic view of organization, signal transmission and A.I. The width of the two paradigms in correlation with the various subject areas shows an estimate of how the relevance of the paradigm is generally considered, although both claim to encompass the entire spectrum.

One of the consequences of this is that concepts of meaning and the objective statistical information concept are defined within two distinct paradigms, making the informational aspect of communication as an objective and quantifiable entity completely independent of any meaningful interpretation from the recipient and any intent from the sender.

The opposition between the two paradigms has another important aspect. It is a confrontation between the scientific and objectivistic realistic views of knowledge and science and the phenomenological-hermeneutic-humanistic approach to meaning, signification and communication. This makes it very difficult to make a unifying theory for LIS that encompasses both the algorithmic way of dealing with intelligence, knowledge and communication in the computer, the social understanding of how meaning is created and evolved in natural languages and finally psychologically how the individual user in front of a document retrieval system actually understands concepts, strings of words, what documents are about and the actual content of documents, since our aim is to organize the retrieval process in a natural way to make the enormous number of documents produced internationally widely available.

The document-mediating system

The main expertise of librarians, archivists and documentalists has always been the storage, indexing, retrieval and mediation of materials carrying data, knowledge, meaning and experience. As a science, its objective is first and foremost to promote communication. This can include recorded measurements and observations, theoretical knowledge, and meanings and visions or experiences, to such media as documents, books, records, tapes, programs, floppy discs, hypertext, compact discs, pictures, films and videograms, from the producer to the user. These mediating forms and future ones can be summarized under the general LIS-concept of a "document" (see e.g. Vickery & Vickery 1987, Buckland 1991). Following Buckland's discussion, I will define a document as a human work with communicative intent towards other living beings that is recorded in a material way.

For librarians and documentalists, information science is primarily concerned with finding the most suitable rules for the design of systems and procedures for collecting, organizing, classifying, indexing, storing, retrieving and mediating those materials which support data, knowledge, meaning and experience. Librarians, documentalists and archivists have done this for thousands of years.

As an offshoot of both indexing and communication to users with different requirements, one must study the origins of the various document types, how they are produced, for which users, and under what economic knowledge domain constraints. It is recognized that producers of documents generally have specific consumers in mind, and these consumers can often be manufacturers themselves. In this way the system closes in upon itself, as Luhmann (1995) underlines for communication systems in general, and then cannot see its surrounding society and culture directly. But it does react to the perturbation and change in the production and use of document types through internal adjustment. This is shown in Figure 2 where the broken arrows represent a structure or result-changing feedback that is

vital for the system's self-organizing ability and its ability to survive through self-adjustment.

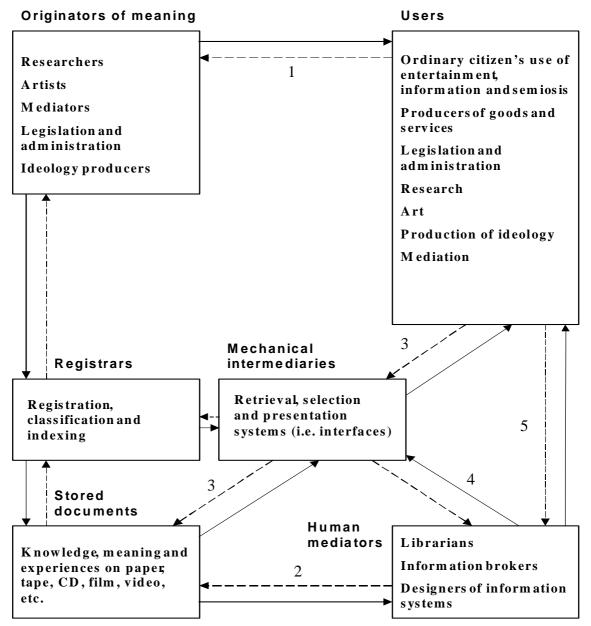


Figure 2: The document-mediating information system as a self-organizing cybersemiotic system with semantic feedback. The unbroken arrows are document transport. The broken arrows are feedback in the form of approval or critique of the contents of documents or of system performance. 1. The direct circulation of documents between producer and user is often seen in the sciences with preprints. 2. The direct access of librarians to a collection. 3. End-user's access directly through on-line systems. 4. The librarian as mediator of the collection through mechanical (electronic) intermediaries. 5. An information broker's mediation of documents to a user.

This is one of the developments – along with the development of cognitive science – that promoted the idea of a unified information science for humans, machines and animals (see for instance Vickery & Vickery 1988). As mentioned previously, the hope of cognitive science is that information-processing will follow certain "universal syntactic, logical and mathematical laws" (Fodor 1987).

One should reflect on the fact that nearly everything, aside from computer programs found on the Internet and in all management information systems, is a document. Therefore this problem is very general and has massive proportions. The first goal is to make intelligent user interfaces. The second is to reorganize databases. The latter does not seem practical or economically feasible for most of the huge international scientific bibliographic databases, since each is built with a rigid scientific classification or thesaurus that controls its indexing practices. Further, they house millions of documents that are already indexed.

The technological impetus for the development of information science

Information science did not really take off until the development of computer technology in the 1960s and its increasing use as information technology in the LIS domain. Seen from society's viewpoint, the problem has primarily been how to handle constructively and cheaply the burgeoning production of documents from science, industry and culture. An industrial aim has been to construct a technology to handle increased access to the buying and selling of knowledge; information is becoming a strategic resource on a level with capital, technology and labor.

The information retrieval industry has become a large-scale industry in the so-called information society. The computer and communication industry has exploded since World War II and is now entering a synthetic phase. Now the computer's various technologies, including calculation technology, telecommunication and language, and more recently its sound and image treatment, are beginning to melt together into a common multimedia interaction technology.

It seems clear that document retrieval, and therefore registration, indexing and classification, plays an unavoidable and growing role in nearly all types of extensive computer networks and all kinds of knowledge-sharing systems. The larger these systems become, the more central the document-mediating component will become and the graver the problem of indexing and intellectual access.

As Blair (1990) has pointed out, there is a qualitative shift in the problem of document retrieval once databases exceed 100.000 documents. First, the number of retrieved documents becomes a problem: there are too many documents to sort for relevance. Second, the level of "noise" becomes intolerable, especially for full text automatic natural language indexed documents. Third, it becomes nearly impossible to estimate recall: nobody knows what of relevance really exists in a base with 15 million documents, such as BIOSIS.

The problem is, on the one hand, that the user is buried in too many documents of different levels of relevance and knowledge quality, and on the other hand, that the user may miss the

most relevant high-quality documents. These are the documents especially suited to the user's problems, interests, knowledge background, focus, knowledge level and time for reading. Anyone who has made a subject search on the Internet understands what I am referring to.

Through the Internet we are improving physical access to electronic documents' information for growing numbers of people, and we are increasing intellectual access for many newcomers and low-level users of document systems. But high-quality intellectual access is becoming a growing problem for those who need it in their daily work, such as researchers, teachers, journalists and managers. Overload, noise, lack of precision and ignorance of recall are modern problems of document retrieval.

One way to improve intellectual access is to create interfaces for users with domain knowledge, but without LIS-technical skills. These difficult cognitive and communicative problems are connected to giving access to users who lack domain knowledge and therefore do not know the specialized meaning of the concepts employed, some of which are the same as those used in everyday language.

There is no doubt that technological development is transforming traditional LIS areas of document retrieval and mediation of knowledge. It is therefore important both theoretically and practically to respond constructively and purposefully to meet this challenge. This requires a scientific basis that encompasses technical, sociological, psychological, and linguistic aspects of the problem of translating peoples' information needs to system-functional queries.

From the last twenty years of computer systems development, it is clear that the manipulation of natural language by machines in a social communicative setting with humans is a major theoretical and practical problem. The meaning of language is one of the pivots around which human existence spins. The problem of how precise meanings of signs and words become fixed within a social and cultural practice is enormous from a traditional mechanistic point of view. Attempts to build interdisciplinary information sciences on the basis of the entropic conceptions of information in information theory and thermodynamics have been theoretically and practically unable to fruitfully deal with the problem of the communication of meaning between humans. The functionalist – or information-processing – paradigm in cognitive science has the additional difficulty of providing a theoretical background for approaching the problem in IR with multi-faceted meanings of words and sentences. As documents are complex semantic sign and language systems, they are some of the most difficult items to handle in computer systems for retrieval by a broad public. As Blair (1990) argues, users are the only reliable source to judge relevance, and only users can turn information into knowledge.

No doubt, development of interfaces that are interactive and graphic could improve search quality for low-level users. But something more fundamental in the organization of our scientific understanding of document-mediating systems is at stake. It is a qualitative difference in the way computers and humans deal with the complexity of information. As Luhmann (1995) points out, then humans reduce complexity through meaning. LIS must

move from a mechanical information-processing understanding of document retrieval based on only cognitive science's information-processing paradigm towards true integration with a more pragmatically semiotic, cybernetic and socio-linguistic theory of understanding in order to improve the design of document-mediating systems. The theoretical foundation of LIS in the IR-area must be replaced by a broader foundation that incorporates the semantic production of meaning. To summarize, LIS has four major problems:

- 1. First the lack of a theory of how to design the best possible document-mediating system for one or more well-known user groups.
- 2. Second the lack of a theory of how to design interfaces for non-specialists for huge document bases originally created for documentalists within certain subject areas or domains, most often scientific and technical such as chemistry, biology and medicine.
- 3. Third the lack of full recognition from computer and software designers and from the arts and sciences of the interdisciplinary complexity and scientific depth of the problems of document mediation.
- 4. Fourth the lack of a full theoretical scientific foundation for the practice of librarianship in the age of the computer. LIS' lack of a fully developed theoretical and scientific self-awareness is a problem for itself, for the recognition by other scientific subjects and research groups of the seriousness of the problems it addresses and the depth of the knowledge it has already acquired through centuries of practice.

In my opinion the information-processing paradigm has never been able to describe the central problems of mediating the semantic content of documents from producer to user that documentalists and librarians deal with. It fails in this regard because it does not address the social and phenomenological aspects of cognition ("becoming informed" in Buckland's 1991 terminology), which is the bottom line of the mediation of documents. This leads to serious doubt about the existence of the scientific object in the form of "objective information processing".

My rejection of the information-processing paradigm of information science is based on views similar to Machlup (1983) and Winograd & Flores' (1987) statements that the original meaning of the word "information" is something a person (or a living system) communicates to another person (or living system). The meaning of information can only be understood by considering living beings in a social and historical context. Furthermore, I agree with Machlup when he suggests that one cannot define information as that which reduces uncertainty. The fact is that some kinds of information will make the receiver more uncertain. But if one knows the social context precisely enough to determine the full spectrum of possible outcomes, then one can use a statistical/entropic information concept as part of the description of the characteristics of a message. I further agree with Searle (1986) that the common link between information-processing in humans and machines is not the fact that both follow rules. Machines behave according to causalities, but only conscious beings can willfully choose to follow rules.

But since the information concept is now firmly rooted in computer informatics and in the information theories of Shannon and Weaver as well of Winerian cybernetics, another strategy would be to abandon the original human communicative meaning of the concept as Shannon's theory of information has never addressed the semantic content of messages. Shannon and Weaver write:

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that they are selected from a set of possible messages.

(Shannon and Weaver: 1969:31-32)

What people and animals treat as "information" is quite different from what Wiener's theory of information suggests. Stonier therefore makes a sensible discrimination between information and meaning, as information to him is objective structure and organization. This is clear, but we must then realize and acknowledge that the theory has almost nothing to do with semantic aspects of cognition or communication between living systems. As von Foerster concludes:

However, when we look more closely at these theories, it becomes transparently clear that they are not really concerned with information but rather with signals and the reliable transmission of signals over unreliable channels (...)

(von Foerster 1980: 20-21)

Thus, information science in the subject area of living systems and humans will not be able to explain vital aspects of the phenomena of cognition and communication, such as meaning and the constraints of social context. It is also well known that to determine the entropy in a system it is necessary to determine in advance what will qualify as macro states and the probability of every state. There is no room for the completely unexpected, and therefore the real creative complexity of nature and language is lost. Thus this approach has other limits on its own level.

In my opinion, knowledge – or knowing (to underline the process) – is a far more complicated "thing" or process than expected by the "information-processing paradigm" described above. According to, for instance, Thomas Kuhn's philosophy of science (1970), nature and human mind are not directly connected. Nature does not speak to us. I would also argue with Maturana and Varela (1980) that nature does not – in the usual meaning of the word – transfer information to us through our observation. We participate through science in a socially, biologically and psychologically influenced interpretation of the world.

Warner (1994) points to the problems for LIS if the meaning of words must be partially inferred from a socio-linguistic context. It is clear that simply matching query words to index words, no matter how sophisticated a partial match and ranking algorithm one has, will always have a low precision because the semantics are not equally well-defined.

In practice, librarians, documentalists and LIS-researchers work every day on the social and practical dynamism of signification and how to relate it to the textual representations in computer systems (not to mention the problems of understanding how the different software system themselves use words). Now through the Internet, corporate Intranets, management information systems, GIS, and file-handling systems, more people are spending more time looking for an exponentially increasing numbers of documents. What LIS needs is an interdisciplinary scientific understanding broad and deep enough to encompass the communicational and organizational aspects of classification and indexing of documents, the computer systems and the producers. Therefore, the theoretical foundation of LIS in the IR-area must be broadened to incorporate the semantic production of meaning in its conception of information science.

In his semiotics, Peirce says that the meaning of a word is its use in society (practice of living). This fits well with the language philosophy of the later Wittgenstein who says that words' meanings are fixed not by definition, but through the "language game" they appear in, such as dialogues, persuasions or seductions. This is what Wittgenstein calls "forms of life," in short, the things humans do. Among these are specific kinds of science. Science is a life form and has its own language games, as document searching does by subject.

The hope of trans-disciplinary theories in information and communication science is that they can deepen our understanding of human to human communication through machine-mediated documents in such a way that we can improve our designs of document-mediating systems.

LIS: The science of document-mediating systems

The subject area of expertise of librarians, archivists and documentalists has always been the storage, indexing, retrieval and mediation of materials carrying data, knowledge, meaning and experience.

- 1. One can therefore define the objective of LIS as first and foremost to promote the communication of desired information from the producer to the user.
- 2. Information can include recorded measurements and observations, theoretical knowledge, meanings, visions and experiences, art and fiction.
- 3. I define data as a given input with a structure that the receiver regards as reliable and usable in a given situation.
- 4. I define a document as a human work with communicative intention recorded in a material way.

I want to clarify Buckland's view that can be construed considering natural things to be documents in themselves. Something culturally and intentionally communicative such as being put into a classification system must be done to natural things before they become documents. But perhaps we can agree by saying that all things are potential documents, just

as everything is potential information. These become documents when they become interesting for members of a communicative knowledge system. But that demands that the object is put into a certain point of view so that it can be viewed from a certain interest, or as Bateson (1973) says: *Information is a difference that makes a difference*.

Data becomes information when it is integrated with a given knowledge process and preunderstanding. It only becomes information when it is received and interpreted by a biopsychological-social knowledge system. The difference between knowledge and information is that information is viewed as a minor part of a knowledge system. But both are dependent on semiotic interpretation if they are to become meaningful.

I agree with Salthe in his view that one cannot consider the meaning of information without an interpretation. We could add to Wiener's statement that (in itself) 'information is information, neither matter nor energy', that *information is also not meaning until it has been interpreted by a living system*.

LIS is concerned with finding suitable rules for the design of systems and procedures for collecting, organizing, classifying, indexing, storing, retrieving and mediating those materials that support data, knowledge, meaning and experience. As an offshoot of both indexing and communication to users with different requirements, one must study the origins of the various document types, how they are produced, for which users they are created and under which economic and knowledge domain constraints they are produced.

It has been realized that producers of documents generally have certain consumers in mind, and these consumers are often part of the group of producers themselves. In this way the system is, as seen by a cybernetician, closed in on it self. The LISA bibliographic database, for instance, is a base of information and library science with documents written by librarians and information scientists, to librarians and information scientists, and mediated by librarians and information scientists.

The cognitive viewpoints opening toward a cybersemiotic concept of information in LIS

In agreement with Ingwersen (1992) one can, as an answer to the humanistic-socially-oriented critique, formulate a broader and less objective concept of information than that of the information-processing paradigm. From the cognitive viewpoint, information is seen as the mental phenomena that documents (consisting of signs and text, depending on the state of knowledge of the recipient) can cause. The examination of these "correct circumstances" is an important part of information science. In connection with the design of information systems for businesses and institutions, one can now speak of information quality. The cognitive viewpoint represents three important developments:

- 1. Information is understood as potential until somebody interprets it.
- 2. The objective carrier of information is a sign.

3. Interpretation is based on the total semantic network, horizons, worldviews and experiences of the person including the emotional and social aspects.

The aim is for the creation of information in the user's mind to be understood as meeting social, cultural or existential needs. This is an important improvement to the intention of cognitive science to create an objective theory about information. One can therefore reformulate information science's aim as follows:

Library and information science devotes itself primarily to the study of systems and methods for classification, indexing, storing, retrieval, and mediation of documents that can cause the creation of information in the user's mind.

The crucial question is that of the interpretation of the document's meaning for the individual in a given organizational or institutional connection, and in a given historical situation. Ingwersen (1996) describes the information need as built from a cognitive state (including previous knowledge), a work task, interest and a domain.

Neither information nor quality is constant phenomena; they change over time. Relevance is the keyword here, and relevance is dependent on the meaning we give to things in relation to our preconceptions. It is these social-pragmatic circumstances that form the context for understanding our informational desires and problems. Ingwersen (1996) successfully develops a matrix with four distinct cognitive forms of information needs relevant for determining search behavior and types of polyrepresentation.

So far, we do not have an explicit theoretical treatment of how varying forms of aboutness come into existence and function in a social context. As information, in this view, develops primarily in an individual mind in front of a document-mediating system, there are no explicit theories about how information develops in social practice. This is where semiotics both as a general and as a social science of meaning generation and interpretation can contribute to the informational view of LIS.

As Blair (1990) has shown, one of the largest problems of subject searching is that indexer and the searcher do not participate in the same language game! Their work and social environments are different and therefore their use of words will be different. This means that their subject descriptions will be different: they will mean different things with the same words, or they will use different words for the same thing. There is no "ultimate" description of a document or of the uses of a subject term. Attempts to make a universal and correct description are pointless because there is no limit to the number of descriptors one could use to characterize a document; semiosis is unlimited (Blair 1990). To limit the spectrum of meanings in a useful way, the indexer must create descriptions in accordance with certain forms of living, such as trading and research, and their language games, also called discourse communities and their discourses.

From this view, it is possible to point to several systematic reasons why the user does not obtain a satisfactory retrieval result and why this situation is so difficult to improve. Although the large domain-specific databases attempt to make the definition of their

classification system (thesauri and controlled keywords, professional indexers) clear and consistent, there is still the human factor of interpretation. The optimum of their actual precision of subject-indexing performance is only 75% consistent between all indexers on a detailed or "deep" indexing of a document. The likelihood that a user would use the subject term in the same way as the indexers – even if the user reads the scope notes carefully – is even smaller, say 60%. This means that the limits of the sets get very fuzzy. Every time one uses a subject term to describe the documents sought, the chance of locating the right one is $0.75 \times 0.60 = 0.45$, even when the system was used to the best of the user's abilities. When one combines four different subject terms to select specific documents, the chance of selecting correctly is then $0.45^4 = 0.04 = 4\%$ (see Blair 1990:106 for an even more merciless example). This problem preexists before using any mechanical model of partial match using probability, weighted factor ranking, vector space, fuzzy sets or hypertext. These are all based on the ideal of good and consistent indexing and a perfect match with the user's application of search terms.

Thus one can only retrieve very small portion of the relevant documents, and a great number of those retrieved will be irrelevant. The more search terms used, the more serious the problem becomes. If one knows the bibliographic base's classification system, there are ways to diminish this problem. But most searchers do not, and even for those who do, their efforts diminish the problem, they do not eliminate it. Information is stored in such a way that only specialists combining subject knowledge with technical retrieval knowledge – investing years of training – really have a chance to retrieve. The system only produces knowledge with the desired precision and scope for certain groups, and then only to a limited extent. There may be physical access for all, but intellectual access only for highly trained specialists.

The basic problem in LIS is that one must perform an intellectual analysis to determine the content of a document in order to achieve precise and useful indexing.

Now, does the indexer determine the objective content of the document? Both practice and hermeneutic theory tell us that the content of a document depends on the context in which it is seen, that is to say, what those who read it know and what their interests are. There are at least three ways of determining the content of the document:

1. The content of the document, as seen from the indexing system (its thesaurus or classification system). In the best cases this is constructed from a profound knowledge of the domain of knowledge in question. This is seldom the case and furthermore, the researcher who wrotes the document probably did not have the classification system and its concepts in mind when writing the paper. The writer might invent some neologisms, a new interdisciplinary subject, or perhaps a whole paradigm opposed to the paradigm underlying the present classification system. Finally, the user often does not share the knowledge background of either the system or the author.

- 2. The content of the document as seen from the author's viewpoint: One can pick words from the text and the indexer can give a description with appropriate words. But the determination of this is an interpretation by the indexer. The main interest of a document retrieval system is that others who need or require the document can find it using their own language game, and this is most likely distinct from those of both the classification system and the indexer.
- 3. The content of the document as seen from the user's viewpoint: The problem here is that in most large document retrieval systems there are so many types of users that the indexer can only index in relation to the largest and most formally well-defined knowledge domains.

As Blair (1990) suggested, one of the major problems of subject searching is that indexers and searchers do not participate in the same language games. Their work and social environments are different, and therefore their uses of words will be different. The hope of trans-disciplinary theories in information and communication science is that they can deepen our understanding of human to human communication through machine-mediated documents in such a way that we can improve our designs of document-mediating systems. To summarize, our major challenge in LIS now is how to map semantic fields of concepts and their signifying contexts into our systems in ways that move beyond the logical and statistical approaches that until now seemed the only realistic strategies given available technology.

Various types and definitions of information

There are various other different aspects of information and meaning that are significant. In his analysis, Buckland (1991: 6) usefully distinguishes the difference between:

- a) Personal knowledge (private, mental),
- b) The process of knowing or becoming informed,
- c) Objective/intersubjective materially registered knowledge (documents) and
- d) Information/data processing, the mechanical manipulation of signals and symbols.

He summarizes this in Figure 3:

INTANGIBLE		TANGIBLE			
ENTITY	Information-as-knowledge Knowledge	1	Information-as-thing Data, document,		2
			recorded knowledge		
PROCESS	Information-as-process Becoming informed	3	Information Processing Data processing, document processing, knowledge engineering	4	

Figure 3: Buckland's matrix of different kinds of information (1991: 6) where he now used the information concept as an overall concept, which will not be our strategy.

I will use similar distinctions: 1. phenomenological knowledge, 2. documents, 3. cognition and 4. information processing. But not use information as an overall concept, but rather as a dualistic concept based on differences that has to be interpreted to generate meaning for an observer.

In my evaluation, the cognitive viewpoint represents an important first step away from the mechanistic information-processing paradigm in cognitive science as a foundation for LIS. It is a step towards a theory that encompasses the social and linguistic complexity of LIS and IR in a more realistic way. Librarians and LIS researchers know this complexity empirically from their own experience, and so far research has modeled different limited aspects of it. But we still have difficulties with the construction of a comprehensive theoretical framework, which can improve consistency in our use of scientific concepts within LIS, guide our research and development of research methods, and finally, provide the background for the interpretation of empirical research. The cognitive viewpoint has made some important changes in the basic view of the communication process in LIS and IR that are compatible with modern semiotics and pragmatic language philosophy. Within the cognitive viewpoint there has been empirical research and developed theory about the situation of the individual user with an information need confronting an information system. But further aspects must be developed for the theory to be comprehensive and broaden it into a general framework for information science.

As in Machlup's theory of information, in the cognitive viewpoint the focus is on the individual. Machlup denies that social systems can communicate. Ingwersen (1996) is open to the study of the influence of knowledge domains on concept formation and interpretation

as launched internationally by Hjørland & Albrechtsen (1995) in the domain analytic paradigm. They give theoretical reasons why classification and indexing should be directed toward the ways signification is created in discourse communities related to different knowledge domains, especially within the different fields of science.

This insight leads to the need for a general semiotic framework of communication and sign processing. We need to open LIS to the results and constructive thinking of a more general theory of how signs – such as words and symbols – acquire their meaning through communication, be it oral or written (Warner 1990). Semiotics should encompass not only social and cultural communication, but also should be able to address natural phenomena such as the communication of biological systems. It should have categories for technical information processing. At the same time, this trans-disciplinary theory should distinguish between physical, biological, mental-psychological and social-linguistic levels, and not reduce them to the same process of information. So far there is not much evidence that a profound and practical understanding of information and communication can be found by reducing them to the mechanical manipulation of symbols. A theoretical understanding of the interpretation of signs by biological-social systems is necessary for speaking about communication of information.

This leads us to the third requirement: a theory of the cognition and communication of signification should be able to encompass different types of systems. Neither the objective syntactic approach of the information-processing paradigm, nor the personal phenomenological approach of Machlup can deliver a framework encompassing communication processes in social, biological and technical systems. We cannot ignore that cybernetic information science that is behind and embedded in the computer is now a general tool for document mediating systems. As Buckland (1991) points out, we must draw on systems theory and cybernetics, and, with Warner and Blair, I add semiotics.

The usefulness of Peirce's approach in LIS

This is all very abstract. Let us therefore consider its usefulness in the Library and information science area to demonstrate what this conceptualization does to help us in one sort of practice; one we have already introduced. The interpreting process according to the semiotic view is unfinishable, just as is scientific knowledge that seeks "the truth". Peirce calls this unlimited semiosis. Signs are woven into meanings, which are linked to societal-cultural communicative praxis and history. Lexical denotations do not define the meaning of signs; these are defined by their use in social life, such as in a language game. Blair points out the significance for LIS for this fundamental understanding of the processes of signification:

In terms of inquiry, the notion of unlimited Semiosis has important consequences for the representation of texts. First of all, there can be no necessary and sufficient (i.e., complete) representation of a text (other than the entire text itself and even this may not be sufficient for retrieval purposes, (...)). Secondly, the standard to be used to judge the usefulness of a particular textual description is not that of "correctness," but one of "appropriateness." In other words, a textual description is neither correct nor incorrect, but, rather, more or less appropriate for a given task and situation.

(Blair 1990: 137-138)

The compatibility between Blair's "appropriateness" and Glasersfeld's "viability" is obvious. Glasersfeld also thinks in terms of different tasks within society, which relates to the "work task/interest" of Ingwersen (1996) and the domains of Hjørland and Albrechtsen (1995). The meaning of words is created through language's cultural-historical background and the social communicative praxis between people, who have their own subjective historical access to the meanings of a sign. People are never in complete agreement about all the meanings of a word or concept. But through the development of customs they may reach an agreement on its meaning within situations experienced jointly. This is significant in various domains of science and the humanities, where long traditions have fixed the meaning of specific concepts. The practice of law has also developed its own special terminology. The pragmatic-semiotic approach is important because it is these connections that constitute the individual's understanding and ability to:

- 1. Decipher the document's signs.
- 2. Decipher the document as a sign in itself.
- 3. To evaluate the relation and value of the sign within the actual situation.

As Blair points out, one must base the organization of document-mediating systems on conventional uses of concepts:

In short, Peirce is pointing out that there can never be a necessary and sufficient explanation or description of the meaning of a sign/expression. In the sense of meaning which we have developed here, this means that there can never be a complete description of the kinds of allowable uses that can be made of a given expression. But this is not a despairing observation; in fact, it puts our analysis into a more thoughtful context. Instead of concerning ourselves with definitive uses of expressions, we can recognize this endless regression of meaning/signification and concentrate on elucidating conventional uses of expressions, realizing that new and creative uses of these expressions are inevitable What is important, then, is not just the uses of an expression, but the conventional uses of that expression in relation to some situation or task at hand.

(Blair 1990: 137)

Peirce is both a phenomenalist and a realist. For Peirce, the meaning of a sign can be boiled down to the habits to which it gives rise and the influence it has on the world (including the inner biological and mental states of the interpretant) now and in the future. His theoretical rhetoric is the science of how signs become effective in a constantly changing historical and social context where there are no final referents. Blair (1990) draws the following conclusions for the understanding of indexing in LIS from this semiotic view of meaning:

In the first place, there is an unlimited number of unique documents which a single subject description can be used to represent. In the second place, there are an unlimited number of subject descriptions that reasonably could be applied to any one document. Traditional indexing theory, though aware of the ambiguity and inconsistency in the assignment of subject descriptions, has never demonstrated a full awareness of the magnitude of this problem, preferring to think of such difficulties as temporary aberrations rather than the first waves of a rising tide of difficulties (...).

(Blair 1990: 169)

This provides a theoretical understanding of the enormous practical problems that have faced classificators and indexers for centuries. The ongoing evolution of signification poses a major difficulty for all document-mediating systems. Every classification system implicitly attempts to define specific meanings of words, and after a few years this becomes a problem for all dynamic knowledge systems. It is an essential for LIS to be able to change classification and indexing systems quickly to follow changes in the meanings of language, while at the same time keeping track of previous records. Since these changes are semantic and related to social practice, we do not yet have a mechanical way to accomplish this. Currently, any document database using words as classification and index terms should have its documents re-indexed every five years to keep in accordance with the present meaning of the words. Furthermore, it would be ideal to have specific classifications and indexing for different user groups with different interpretations of keywords, to account for their different types of educations, sciences, and practices.

Some would argue that Peirce's semiotics do not tell us much about texts and language. But as Blair claims, Peirce is in general agreement with Wittgenstein, who in his "Philosophical Investigations" (1958) Sect. 43 says that: "For a *large* class of cases – though not for all – in which we employ the word "meaning", it can be defined thus: the meaning of a word is its use in the language. Suggests that the meaning of a word is equivalent to its use within a specific "language game" within a "life form" (*Lebensform*) as mentioned earlier? Language-games, forms of life and rule-following are shaping the meaning of any word. It is a matter of what we do with our language in social practice. It is not something hidden inside anyone's mind or brain. Words, gestures, expressions come alive only within a language game, a culture, a form of life. If a picture for example means something then its meaning is not an objective property of the picture in the way that its size and shape are; it means something to somebody. The same goes for any mental picture. Mankind lives in cultural communities or forms of life, which are self-sustaining, self-legitimating, logically and normatively final. Actually Wittgenstein also mentioned being a Lion as a life form, when he explained that we could not understand it if it could talk.

Blair's work attempts to integrate the crucial insights from Peirce's semiotics with Wittgenstein's pragmatic language philosophy in order to re-examine the problems of IR and LIS in a new light. Blair argues that the semantic socio-pragmatic basis for meaning is a fundamental aspect of Peirce's "unlimited semiosis." Blair demonstrates how essential it is for LIS to realize that to comprehend a concept's meaning, indexers and classifiers must understand its use for a given producer, in a given specific knowledge domain, and for a given user group. I suggest that this fundamental semiotic and socio-linguistic knowledge is the theoretical foundation behind the domain analysis of Hjørland and Albrechtsen (1995) as well as the cognitive viewpoint that employs concepts of "about-ness".

But understanding the language game of the knowledge domain from which a document originates is not enough. One must also understand the language game of which the IR-process is a part. One aspect of this is the knowledge domain from which the user comes, but just as important are the intentions and social expectations users have of the system, as determined by their own understandings of their tasks. Blair writes:

Various kinds of activities (Forms of Life) can serve as a context for the retrieval of subject material. Activities such as defending or prosecuting a lawsuit, patent searching, conducting research, making a business decision, etc., all may make use of subject searching at various times and at various levels of intensity. The nature of the activity being pursued influences subject searching in two important ways: In the first place, the language of the activity, its jargon or cant, will determine which words will be used to describe and ask for subjects. Some activities have or use information that breaks down readily into subject areas, such as academic disciplines (especially the "harder" or more formal ones), while others have and use information that may not be as readily classifiable (think of activities that deal with new or innovative products or processes, such as new marketing, engineering or medical techniques, to name only a few). The other way in which the nature of an activity can influence retrieval is in determining the level of exhaustivity needed for satisfactory retrieval. Patent searching, the defense of a lawsuit, or searching to support original research all demand that the information retrieval which supports their activity be as exhaustive as possible. The activities of "Just keeping informed", browsing, or introducing oneself to a new field require less exhaustive searches to be conducted.

(Blair 1990: 158)

As we shall see, even more language games are actually in play in the IR-process, as is already understood by the cognitive viewpoint and its multiple uses of the concept of "about-ness" (Ingwersen 1992). In my opinion, the concept of a language game provides both a theoretical and pragmatic framework for understanding about-ness and provides an important link between this idea and broader social-pragmatic theories of language and cognition that are of importance for LIS.

Language games are not only connected to users' searches for documents, but also with the overall design and maintenance of the system and the intentions behind the production of documents. Figure 4 illustrates some of the language game systems involved in IR for a document-mediating system.

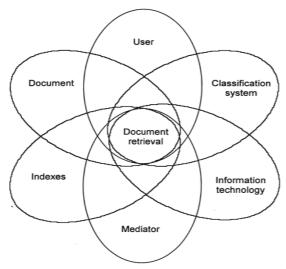


Figure 4: The different language games involved in IR for a document-mediating system. A substantial part of the challenges facing LIS. are not so much technical difficulties as socio-linguistic difficulties of making concepts communicate within and across different types of language games. The user has his own language game, but also must handle documents with a language game of their own determined by their authors, ordered by the language game of a classification system, indexed according to the language game of the

indexer, and searched for in the language game of the search language. If the user has someone to help, then this problem will be further interpreted through the language game of the mediator (often that of the librarian or documentalist).

The skill of the librarian to mediate between different language games in document mediation is a complicated one that becomes increasingly complex every year. Because the number of documents and users grows exponentially, economic considerations force us to seek automated solutions to these problems. It is very important to research how to achieve the most productive integration between machine and human skills. But the economic necessity of using machines should not obscure the fact that the central challenges facing LIS are often the interface between the socio-pragmatic linguistic and the logic or algorithmic processing of sign vehicles that computers perform.

Indexing is central for information storage and retrieval. The indexer allows a descriptor to represent something else – such as a document – so that it can be found by means of something else, on the basis that these entities in some respect share the same content/idea. This description parallels the description of the sign.

It is necessary to identify the sign relations between descriptors and documents to understand these relations as signs, and to recognize that the signs can alter sign categories – in other words, the sign alters nature according to the person who interprets it. The relation between the descriptor and the document is essential for understanding and recognizing what happens during indexing and retrieval. Through a semiotic discussion of this relationship is it possible to describe the nature of descriptors from the users' level of knowledge. We have included the user in the descriptor-document relation. However, we have seen that there are also other semiotic webs involved in the user-descriptor-document relationship – namely the semiotic webs of the indexers and the semiotic webs of the authors. If information retrieval is to be successful, these semiotic webs must approximate each other.

I am sympathetic to van Rijsbergen and other colleagues' attempts to resolve the difficulties of computer document mediation by creating new kinds of logics, although I would still argue that the core relationship for mediating documents is semantic, and that semantic relationships are not built primarily on logic, but rather on motivated relations influencing the intentionality of conscious awareness. These are established through the evolution of living systems (ethology) and through the history of life forms and language games in societies. They are created as structural couplings of significance within a semiotic web (such as Peirce's triadic semiotics), established through the living system's relationships to nature and other bodies within social systems.

The computer has seduced us into framing our questions within its algorithms. It seems we have forgotten to maintain and develop a theoretical framework for our subject area that allows us to see beyond the horizon of the computer and to make demands of those researchers developing computer systems. If we do not provide a metatheoretical description of our own area, it becomes difficult for others, such as computer scientists and software developers, to understand that they have entered a new territory with different rules. We

must provide a strong theoretical understanding of the difference between physical and intellectual access. The growth of the Internet makes this knowledge more important every day.

One should start with a pragmatic analysis of the informal communication system. This is the most powerful semiotic force to which any information system must adapt, and as Lakoff (1987) demonstrates, its semantic patterns are neither logical nor random – they are motivated. This accords with the cybernetic view of information as generated within an autopoietic system, and language communication as occurring within generalized media. Motivation stems from the type of media, but the actual language game chosen within the media determines a large part of the motivation for the relationship between concepts. If there is no proper feedback between producers, indexers, and users, the system will not produce information – it will not fulfill our expectations. We all participate in several language games simultaneously, but professionally we must consciously select and maintain one at a time whenever possible. As information is only potential when there is no interpretant, the only information in our systems is relevant retrieved documents. This further supports much of Bates' work on the sense-making approach (1989).

The pragmatic approach generally means, as previously mentioned, that a philosophy of science analysis of the domains/subject area/work tasks and paradigms in science, as well as a knowledge sociological analysis of communication patterns such as the discourse analysis of written text, is important for describing the decisive context of the use of our systems. They must be adjusted to our context, work task, and the budget allotted the research. These methods should be supplemented by questionnaires, association tests, and registration methods. The expenses of this research are a challenge, but the willingness to pay for basic research is connected to the users' awareness of how central insights into the socio-pragmatic linguistic framework are to the performance of the designed systems. We are moving past the phase of unreflective fascination with electronic systems and into a more realistic evaluation of how they can help us mediate communications between humans via documents. If one considers Ingwersen's (1992) analysis of what a mediator system must do to function properly, one realizes we cannot expect machines to solve the complexity of human communication without human mediation.

This knowledge also tells us that there is limited utility to the non-professional users of the enormous scientific and technical bibliographic bases where many millions of documents have been categorized into Boolean systems by trained documentalists. Here, the users are the documentalists themselves, and the trained researchers from part of the domain, search bases that have not been made generally accessible through the Internet. New digital libraries based on the same outdated principles and word-to-word matches are constantly being established also in the industry. A bibliographic system such as BIOSIS, based on the present theory, will only truly function within a community of biologists. This means that both the producers and the users must be biologists – and so must the indexers. Even then there will be difficulties, because the producers and the users of the bibliographic database will also be researchers. This is a life form that follows a language game different from that of indexers. But if indexers maintain contact with both users and producers, solicit their feedback, attend their conferences, and investigate their ways of utilizing literature and

scientific concepts, the system will holistically produce information. One should not understand document-mediating systems as merely information keepers and deliverers. They are information producers, once we include interactions with users as part of the system!

In the enormous domain-specific systems, we have to accept a centrally organized knowledge system. We can simplify through menu-driven systems only at the cost of speed and precision. We can help users understand what kind of system they are working with by providing thesauri to consult and work from directly. We can remind them to consider specific vital details by asking them to answer questions as part of an obligatory procedure. All this is now done in new types of interfaces. Blair (1990) suggests offering users the opportunity to view extracts of papers that the use of specific index terms will access, and what other users have accessed using similar searches. Any technique that helps users understand the language game they are participating in, how it is structured, and how words work within it is fruitful when combined with opportunities to navigate, explore, and learn the system by oneself.

In these cases we cannot bring the system to the user, so we must bring the user to the system. This will not happen if we simply install a natural language processing interface that pretend to the users that this system will do most of the thinking for them. We should clarify that these systems only help users who do not have the time or ability for other types of search process, because users will have practically no control over the processes by which papers are accessed. This might nevertheless be useful if these users want only a few documents on a subject of interest. The same can be said for the automatic indexing of full-text documents (Blair 1990), unless it is in a sharply delineated and rigidly formalized subject area. Automated procedures give users little insight into what occurs within a system. Users have very little opportunity to control the language game they are participating in. This does not even broach the issues that arise when index terms from one language game are used to seek documents in another.

The problem of intellectual access cannot be resolved by intelligent user-interfaces in the pre-existing Boolean system. Nor will the addition of automated indexing, including natural or knowledge-domain specific language manipulation, or including full text systems (Blair 1990). Undoubtedly each is useful within limited contexts. In currently existing, large scientific bibliographic databases, considerable efforts have been made to deliver interfaces that obligate users to pay attention how the base is structured and remember its most relevant aspects. By reading manuals, one can acquire a simplified theoretical impression of how the controlled index terms are used. Blair (1990) suggests that users should have the opportunity to view samples of papers that are retrieved using a search term, so that users can gain experience about how words function within the language game of the classification system, and through this learn their meanings. The BIOSIS Previews manual, for example, gives theoretical examples of this kind. It is also important to allow as much opportunity for exploring as possible.

When we contemplate designing a new document-mediating system from the bottom up, the suggestion is to specialize document-mediating systems for specific knowledge domains,

knowledge levels, and points of interest, and to consider the size of the system. This means constructing bases entirely from users' needs and conceptual worlds. We must supplement current methods with pragmatic analysis of discourse communities with various knowledge domains, both scientific and non-scientific.

Most current bibliographic databases contain documents produced by different paradigms, specialties, and subject areas, all of which have different language games even when they share a vocabulary. I only need to mention how data-engineers, cognitive psychologists, and information scientists use the concept of information, or how Newtonian physics and Einstein's general relativity use the concept of space. Each subject area with interest in the documents of a database should have these documents indexed according to their own language game in order to make precise searches possible. As is already acknowledged in BIOSIS, for example, chemists, physicians, and biologists each have specific terms for chemicals, illnesses, and classifications of plants and animals that are respected by the BIOSIS indexing procedure. But under current indexes, as a biologist I must use chemical notation searching for a chemical, and chemists must use the correct biological name for a plant to find articles about a chemical substance it produces. Not addressed are those words common to all three subject areas, but that has different meanings because they are part of different language games. We must develop methods to more fully analyze the discourse communities in various knowledge domains, scientific and non-scientific, theoretical and practical. We must get a firmer grasp on the social-pragmatic connotations of words and concepts in order to integrate them into the semantics of semiotic nets as a basis for thesaurus building.

As a result, one of the large research areas of LIS is how to integrate bibliographic databases and full-text databases into different domains, organizations, interests, and levels in organization. This demands that one to distinguish and characterize different domains, levels, and language games in, for instance, an organization. In addition to the methods already employed by LIS, these analyses will benefit from methods derived from discourse and conversation analysis, as well as from socio- and ethnolinguistic empirical analysis of cultural communication.

Bøgh Andersen (1990) delivers a semiotic framework for a computer semiotics and some applicable methods from a non-LIS context. Most fields today are, at least to some degree, interdisciplinary – BIOSIS is a good example, as it is relevant to medicine, chemistry, and the behavioral sciences – and one could for imagine that eventually interest groups from different domains would develop their own systems for indexing documents, so they can chose their own point of entry to these systems. In addition, there will be various offers to visualize systems and their language games aimed at searchers who lack domain knowledge or technical search knowledge, combined with many possibilities for navigation. But that is a far cry from what users want.

What is it many of us want? Well I want the document based to be like the way I organize the books in my office. I would like to be able to put in my electronic knowledge profile, build up over my entire life. When I load that into the system I want all the books re-indexed

according to the way I define the groups and index words according to my understanding and use of the concepts.

Further, I think most of us do not put them in alphabetic order after writer or title in our offices if we have more than 100 books there, but tend to group them according to various subject areas, which are again more defined from the way we use them – for instance in teaching various courses or participating in various research groups or projects – than from some international definition of the area. Thus the ultimate wish of a user of a database is that all documents with only a moments notice are related to his or hers own personal knowledge profile. The wish of the database designer has always been to construct the perfect universal system that all users had to learn. Thus as in the sciences and as in the discussion about rationality in our culture we are torn between those striving towards the universal truth and those who strive for meaning be it individually, culturally or even universal. One of our civilizations great works on that is Gadamer, *Truth and Method*, which develops the hermeneutic view in a reflection also on the truth and method concepts in the sciences. Peirce's semiotics can be seen as a transdisciplinary paradigm that also includes phenomenology and hermeneutics.

References

Barbieri, M.(2001): The organic Codes: The birth of Semantic Biology, PeQuod.

Bates, M. J. (1989): The design of browsing and berry-picking techniques for the online search interface, *Online Review*, Vol. 13, No. 5, pp. 407-424.

Bateson, G. (1973): Steps to an ecology of mind, Paladin, USA, Great Britain.

Blair, D.C. (1990): Language and representation in information retrieval. Amsterdam: Elsevier.

Buckland, M. (1991): Information and Information Systems, New York; London: Greenwood.

Bøgh Andersen, P. (1990): A theory of computer semiotics: semiotic approaches to construction and assessment to computer systems. Cambridge: Cambridge University Press.

Fodor, J. A. (1987): Psychosemantics: the Problems of Meaning, Cambridge, MA: MIT Press.

Foerster, H. von (1980): Epistemology of communication, in Woodward, K (ed): The Myth of Information: Technology and Postindustrial Culture. London: Routledge & Kegan Paul.

Gadamer, H.-G. (1975): Truth and Method, New York: Seabury Press.

Hayles, N.K. (1999): How we became posthuman: Virtual bodies in cybernetics, literature, and informatics, The University of Chicago press, Chicago and London.

Hjørland, B. & Albrechtsen, H. (1995): Toward A New Horizon in Information Science: domain Analysis, Journal of the American Society for Information *Science*. Vol. 46, No. 6. pp. 400-425.

Hoffmeyer, J. and Emmeche, C. (1991): Code-Duality and the Semiotics of Nature, in M. Anderson and F. Merrell (ed): On Scientific Modeling, New York: De Gruyter, pp. 117-166.

Ingwersen, P. (1992): Information Retrieval Interaction. London: Taylor Graham, pp. 246.

Ingwersen, P. (1996): Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory, Journal of Documentation, Vol. 52, No. 1, March 1996, pp. 3-50.

Kuhn, T. (1970): The Structure of Scientific Revolutions, 2nd enlarged ed. Chicago: The University of Chicago Press.

Lakoff, G. (1987): Women, Fire and Dangerous Things: What Categories Reveal about the Mind, Chicago and London: The University of Chicago Press.

Luhmann, N. (1995): Social Systems. Stanford, CA: Stanford University Press.

Machlup, F. (1983): Semantic Quirks in Studies of Information, in Machlup, F. & Mansfield, U.: The study of information: interdisciplinary messages, New York: John Wiley & Sons. pp. 641-671.

Maturana, H & Varela, F. (1980). Autopoiesis and Cognition: The realization of the Living, London: Reidel.

Nöth, W. (1995): Handbook of Semiotics. Bloomington and Indianapolis: Indiana University Press.

Peirce, C.S. (1994 [1866-1913]): The Collected Papers of Charles Sanders Peirce. Electronic edition reproducing Vols. I-VI ed. Charles Hartshorne & Paul Weiss (Cambridge: Harvard University Press, 1931-1935), Vols. VII-VIII ed. Arthur W. Burks (same publisher, 1958). Charlottesville: Intelex Corporation.

Prigogine, I. & Stengers, I. (1984). Order Out of Chaos: Man's New Dialogue with Nature, New York: Bantam Books.

Schrödinger, E. (1967). What is life? The physical aspect of the living cell and mind and matter, Cambridge: Cambridge University Press.

Searle, J. (1989): Minds, Brains and Science. London: Penguin Books.

Shannon, C. E. & Weaver, W. (1969): The Mathematical Theory of Communication, Urbana: The University of Illinois Press.

Warner, J. (1990): Semiotics, Information Science and Computers, Journal of Documentation, Vol. 46, No.1, pp. 16-32.

Warner, J. (1994): From Writing to Computers. London: Routledge.

Vickery, A. & Vickery, B. (1988): Information Science - Theory and Practice, London: Bowker-Saur.

Winograd, T. & Flores, F (1986): Understanding computers and cognition: a new foundation for design, Norwood, N.J.: Ablex Pub. Corp.