ARIST 41 2007

Semantics and Knowledge Organization

Birger Hjørland

Introduction: The importance of semantics for information science (IS)

The aim of this chapter is to demonstrate that semantic issues underline all research questions within *Library and Information Science* (LIS) (or just IS[1]) and in particular the subfield known as Knowledge Organization (KO). Further the aim is to demonstrate that semantics is a field influenced by conflicting views, why it is important to argue for the most fruitful one. Finally the chapter demonstrates that LIS so far have not addressed semantic problems in any systematic way, why the field is very fragmented and without a proper theoretical basis. This chapter is a review that focuses on broad interdisciplinary issues and the long term perspective.

The theoretical problems involving semantics and concepts are very complicated why this paper starts by considering tools developed in KO for information retrieval (IR) as basically semantic tools and thus establishing a specific IS focus on the relation between KO and semantics.

It is well known that thesauri consist of a selection of concepts supplemented with information about their semantic relations (such as generic relations or "associative relations"). Some words in thesauri are "preferred terms" (= descriptors) others are "lead-in terms". The descriptors represent concepts. The difference between "a word" and "a concept" being that different words may have the same meaning and similar words may have different meanings, whereas one concept expresses one meaning.

For example has the word "letter" according to WorldNet 2.1 five senses, among them: 1) a written message addressed to a person or organization and 2) a letter of the alphabet, alphabetic character. In a thesaurus such meanings are distinguished, e.g. by parenthetical qualifiers, as done in *Thesaurus of ERIC Descriptors* (2001)**:**

Letters (Alphabet);
Letters (Correspondence);

By means of Use/Used for relations the thesaurus manages the synonymy relations. By means of parenthetical qualifiers the thesaurus manages the homonymy relations. By means of semantic relations between descriptors (concepts) such as NT, BT, RT, the thesaurus establishes a structure of a subject field:

> "Most thesauri establish a controlled vocabulary, a standardized terminology, in which each concept is represented by one term, a descriptor, that is used in indexing and can thus be used with confidence in searching; in such a system the thesaurus must support the indexer in identifying all descriptors that should be assigned to a document in light of the questions that are likely to be asked. . . .A good thesaurus provides, through its hierarchy augmented by associative relationships between concepts, a semantic road map for searchers and indexers and anybody else interested in an orderly grasp of a subject field." (Soergel, 2004).

It should now be clear that a thesaurus is basically a semantic tool because "the road map" it provides is semantic: the relations shown between the concepts in a thesaurus are semantic relations.

[1] LIS and IS are regarded as synonyms in this chapter. Other researchers do not regard them as synonyms. This example of semantic relations is in itself part of the problems that KO faces. People who claim that they are not synonyms should be able to say whether a given paper belong to IS or to LIS.

What is the case with thesauri is more or less the case with all kinds of what Hodge (2000) presents as “knowledge organizing systems” (KOS) in the following taxonomy:

Term Lists

- *Authority Files*
- *Glossaries*
- *Dictionaries*
- *Gazetteers*

Classifications and Categories

- *Subject Headings*
- *Classification Schemes*
- *Taxonomies*
- *Categorization Schemes*

Relationship Lists

- *Thesauri*
- *Semantic Networks*
- *Ontologies*

All these items discussed as KOS by Hodge represents selections of concepts more or less enriched with information about their semantic relations. Semantic networks, for example, are kinds of KOS utilizing more varied kinds of semantic relations compared to thesauri (while authority files are kinds of KOS displaying only poor information about semantic relations). Because those systems are all basically about concepts and semantic relations, important knowledge about concepts and semantics should be important for research and use of any of those systems, and different semantic theories must imply different principles of knowledge organization. In other words: Researchers in KO should base their work on a fruitful theory of semantics. This kind of basic research has, however, been almost absent in LIS.

We have now argued that what have been termed KOS by Hodge may all be considered semantic tools. We will now have a closer look at and a discussion of the term “knowledge organizing systems".

There are kinds of KOS which Hodge (2000) does not consider.

Hodge does not, for example consider bibliometric maps such as those provided by, for example, White & McCain (1998). In such maps citation patterns may be displayed by authors and/or by terms (e.g. from descriptors). Such maps are thus displaying a certain kind of semantic relations based on citing behavior (and the relation between terms on such a map displays a certain kind of semantic distance). Bibliometrics is important to include in the concept of KOS, both because of theoretical and practical reasons.

There are other kinds of KOS that Hodge (2000) do not consider. It could be argued that, for example, encyclopedias, libraries, bibliographical databases and many other concepts used within LIS should be considered kinds of KOS. Also concepts outside LIS such as the system of scientific disciplines or the social division of labor in society are, for example, very fundamental kinds of KOS. KOS in a narrow LIS oriented sense are the systems related to organizing bibliographical records (in databases). KOS in a wide sense is related to the organization of literatures, traditions, disciplines and people in different cultures. It will be argued that KOS in the wide sense are important to consider also for narrow LIS concerns.

While all KOS considered by Hodge, in addition to other kinds such a bibliometric maps may be considered semantic tools, not all kinds of KOS are. The system of scientific disciplines, for example, is not a semantic tool. The term “semantic tools” should be preferred for systems which

provide selections of concepts more or less enriched with information about semantic relations, while KOS should be used as a broader term including, but not limited to semantic tools.

The field of Knowledge Organization within LIS is thus concerned with the construction, use and evaluation of semantic tools for IR. This insight brings semantics to the forefront of LIS. This view is shared with Khoo & Na (2005), who declare that the study of "semantic relations is the new frontier for information science in the 21st century".

Given that concepts are the meaning behind words and that semantics is the study of meaning the study of concepts, meaning and semantics should form one interdisciplinary subject field. Today it is, however, very scattered and difficult (covering, among other fields philosophy, linguistics, psychology and cognitive science, sociology, computer science and information science, IS). In addition to the disciplinary scattering of research in semantics, the field is based on different epistemological assumptions with roots going hundreds of years back in the history of philosophy. Moreover, the field seems theoretically muddled.

Semantics is, by the way, not just about word meaning. Pictures as well as other signs are also the objects of semantics. The way semantics is viewed and discussed in this chapter may, by many people, look more like semiotics (the study of signs in general) than like the way semantics is often understood. The relation between semantics and semiotics is itself a controversial issue. The focus on semantics rather than on semiotics in this chapter is motivated by the fact that thesaural relations (like KOS in general) are semantic relations as discussed above.

The status of semantic research in information science

Van Rijsbergen (1986, p. 194) pointed out that the concept of meaning has been overlooked in IS, why the whole area is in a crisis. The fundamental basis of all the previous work – including his own – is wrong, he claims, because it has been based on the assumption that a formal notion of meaning is not required to solve the information retrieval (IR) problems. This statement by a leading researcher should justify a closer cooperation between IS and the multidisciplinary research done in semantics. Few researchers have, however, met this challenge and not much consideration has been done concerning the nature of semantics and its implication for IS, although some beginnings are made.

Among the presentations of semantic issues in knowledge organization and IS are Bean & Green, 2001, Beghtol, 1986, Blair, 1990 & 2003, Bonnevie, 2001, Brooks, 1995 & 1998, Budd, 2004, Dahlberg, 1978 & 1995, Daily, 1979, Doerr, 2001, Foskett, 1977, Frohmann, 1983, Green; Bean & Myaeng, 2002, Hammerwohner & Kuhlen, 1994, Hedlund, Pirkola & Kalervo, 2001, Hjørland, 1997& 1998, Khoo & Na, 2005, Qin, 1999 & 2000, Read 1973, Song & Galardi, 2001, Stokolova, 1976, 1977a+b and Vickery & Vickery, 1987.

These contributions are very different and difficult to present in any coherent way because they are not related to each other or systematically related to broader views. Some of them try to base their view on an explicit philosophy (e.g. on "Activity Theory" (Hjørland, 1997) or on Wittgenstein's philosophy (Blair, 1990 & 2003; Frohmann, 1983); others, e.g., Vickery & Vickery (1987) base their view on cognitive psychology, while many just present their own common sense views without trying to relate to general theories (e.g., Foskett, 1977). A book such as Green, Bean & Myaeng (2002) should be praised for its attempt to present an interdisciplinary perspective. Both this book and reviews such as Khoo & Na (2005) fail however to consider much previous research within information science (such as many of the references listed above) and thus to provide a

historical perspective on the relation between semantics and LIS. They also fail to provide a discussion of basic issues in semantics and thus to argue systematically for a specific theoretical view. This state-of-the-art leaves us without a clear line of progress. Without proper theoretical frames of reference, empirical research becomes fragmented and almost impossible to overview.

Much research is also based on technicalities without much concern with basic semantic issues. This is the case with the bibliometric research about semantic relationships between highly cited articles (e.g., Song & Galardi, 2001), in the technique known as "latent semantic indexing" or "latent semantic analysis" (e.g., Ding, 2005; Dumais, 2004) and of course in particular the new concept considered by many the most important frontier in knowledge organization: "the semantic web" (Antoniou & van Harmelen, 2004. Berners-Lee et al., 2001; Fensel, et al., 2003). All such technologies are providing semantic tools, why different view in semantics should make an important difference for how such technologies should be evaluated.

There are also papers (such as Budd, 2004) which introduces important philosophical and semantic views in LIS, but which are not specific in their implications for knowledge organization. There is a danger that the philosophical insights remain too isolated and too vague.

The question concerning the relation semantics and KO may be turned upside down: We may ask from which theoretical perspectives KO has been approached? Which views of semantics have been implied by those approaches?

KO has a long tradition within LIS. Among the classics in the field is Bliss (1929). In order to discuss the relations between semantics and KO we may ask: What approaches have been used in the field of KO during its history? How do they relate to semantic theory? Broughton et al. (2005) suggested that the following traditions in KO are most important to consider:

1. The traditional approach to KOS expressed by classification systems used in libraries and databases, including DDC, LCC and UDC
2. The facet-analytical approach founded by Ranganathan
3. The information retrieval tradition (IR)
4. User oriented / cognitive views
5. Bibliometric approaches
6. The domain analytic approach
7. Other approaches. Many other approaches have been suggested. Among them semiotic approaches, "critical-hermeneutical" approaches, discourse-analytic approaches and genre-based approaches. An important trend is also an emphasis on document representations, document typology and description, mark up languages, document architectures etc.

Given that KOS essentially are semantic tools should different approaches to KO reflect different approaches to semantics. This connection can only be answered briefly here. The traditional approach to classification introduced the principle of literary warrant and thus based the semantic relations in the scientific and scholarly literature. This was (and is) often done on positivist premises: The scientific literature is seen as representing facts about knowledge and structures in knowledge and that subject specialists are able to make true and objective representations of in KO (thus tending to neglect conflicting evidence and theories). The facet analytic approach tends to base KO more on a priory semantic relations. Its methodology is more based on the application of (logical) principles than on the study of evidence in literatures (although this is also to some degrees visible in the tradition). The IR tradition sees the semantic relations as statistical relations between signs and documents. It is atomist in the sense that it does not consider how traditions, theories and

discourse communities have formed the statistical patterns it observes. User-oriented and cognitive views tend to replace literary warrant with empirical user studies and thus to base semantic relations on users rather than on the scientific literature. The bibliometric approach considers documents to be semantically related if they cite each other, are being co-cited or bibliographic coupled. Again are the semantic relations based on some kind of literary warrant, but in a quite different way compared to the traditional approach. The domain-analytic approach is rather traditional in its identification of semantic relations based on literary warrant. It is not positivist, however. It regards semantic relations as determined by theories and epistemologies, which more or less influence all fields of knowledge. Many recent approaches to KO, including semiotic and hermeneutic approaches may be regarded as related to the domain-analytic approach.

What is indicated above is that different approaches to KO imply different views on semantics. This is, however, a point that has not been considered in the literature before.

Semantics and the philosophy of science

Different theories and epistemologies are more or less conflicting and may be more or less fruitful (or harmful) for information science. It is important to realize this and to take the risk defending a particular theory. If this is not done the views will never be sufficiently falsified, confirmed or clarified. In the process of defending a particular view, one has to find out, what other views are consequently rejected. As the pragmatic philosophers suggest: In order to make our thoughts clear we shall ask: What practical consequence does it make whether one or another view is taken as true? (Or whether one or another meaning is taken as true?). If no practical implications follow, our theory (or meaning) is of no consequence and thus not important.

Peregrin (2004) suggests that the two main paradigms in semantics are the one developed by logical positivists such as Rudolph Carnap (and the young Wittgenstein) on the one hand and the one developed by pragmatic philosophers such as John Dewey (and related to, among others, the late Wittgenstein) on the other hand. The positivist semantics suggests that expressions 'stand for' entities and their meanings are the entities stood for by them. The pragmatic semantics suggests that expressions are tools for interaction and their meanings are their functions within the interaction, their aptitudes to serve it in their distinctive ways[2].

This dichotomy is also used by Hjørland & Nissen Pedersen (2005) about the foundation of a theory of classification for information retrieval. Their arguments may be summarized as follows:

1. Classification is the ordering of objects (or processes or ideas, whatsoever) into classes on the basis of some properties. (The same is the case when terms are defined: it is determined what objects fall under the term).

2. The properties of objects are not just "given" but are only available to us on the basis of some descriptions and pre-understandings of those objects.

3. Description (or every other kind of representation) of objects is both a reflection of the thing described and of the subject doing the description. Descriptions are more or less purposeful and theory-laden. Pharmacologists, for example, in their description of chemicals, emphasize the medical effects of chemicals, whereas "pure" chemists emphasis other things such as their structural properties.

[2] In the sociology of science is the debate between "meaning finitism" and "meaning determinism" a related theoretical discussion. (Cf., Barnes, 2002; Bloor, 1997, pp. 1-3 & 9-11; Haukioja, 2005; Klaes, 2002; Larsson, 2003 and Weber, 2005). An important critique of semantic assumptions generally made in science is Harris (2005).

4. The selection of the properties of the objects to be classified must reflect the purpose of the classification. There is no "neutral" or "objective" way to select properties for classification because any choice facilitates some use while at the same time limits other use.

5. The (false) belief that there exist objective criteria for classification may be termed "empiricism" or "positivism", while the belief that classifications are always reflecting a purpose may be termed "pragmatism". The paper is thus an argument for the pragmatist way of understanding.

6. Different domains (chemistry and pharmacology) may need different descriptions and classification of objects to serve their specific purpose in the social division of labor in society. The criteria for classification are thus generally domain-specific. Different domains develop specific languages (LSPs) that are useful to describe, differentiate and classify objects in their respective domain.

7. In every domain different theories, approaches, interests or "paradigms" exist, which also tend to describe and classify the objects according to their respective views and goals.

8. Any given classification or definition will always be a reflection of a certain view or approach to the objects being classified. Ørom (2003), for example, shows how different library classifications are reflecting different views of the Arts. Ereshefsky (2000) argues that Linnaean classification is based on criteria that are pre-Darwian and thus problematic. Sometimes, however, a given classification seems to be immune to criticism. This may be the case with *the Periodical System* of Chemistry and Physics. Such immunity is caused by a strong consensus in the underlying theory.

9. A given literature to be classified is always - more or less - a merging of different domains and approaches/theories/views. Such different views may be explicit or implicit. If they are implicit they can be uncovered by theoretical and philosophical analysis.

10. Classifications and semantic systems that do not consider the different goals and interest reflected in the literature of a given domain are "positivist". The criteria for classification should be based on an understanding of the specific goals, values and interest at play. They are not to be established a priory, but by "literary warrant": by examining the literature. This cannot either be done in a "neutral" or "objective" way, but may be done more or less qualified by considering the different arguments.

In her reply Sparck Jones (2005) acknowledges the pragmatic point of view. Her final suggestion is, however:

> "At the same time, one of the most important techniques developed in retrieval research and very prominent in recent work, namely relevance feedback, raises a more fundamental question. This is whether classification in the conventional, explicit sense, is really needed for retrieval in many, or most, cases, or whether classification in the general (i.e. default) retrieval context has a quite other interpretation. Relevance feedback simply exploits term distribution information along with relevance judgements on viewed documents in order to modify queries. In doing this it is forming and using an implicit term classification for a particular user situation. As classification the process is indirect and minimal. It indeed depends on what properties are chosen as the basic data features, e.g. simple terms and, through weighting, on the values they can take; but beyond that it assumes very little from the point of view of classification. It is possible to argue that for at least the core retrieval requirement, giving a user more of what they like, it is fine. Yet it is certainly not a big deal as classification per se: in fact most of the mileage comes from weighting. And how large that mileage can be is what

> retrieval research in the many experiments done in the last decade have demonstrated, and web engines have taken on board." (Sparck Jones, 2005).

We agree that meanings and classification criteria are implicit in the literature to be retrieved, as outlined above. Spark Jones asks "whether classification in the conventional, explicit sense, is really needed for retrieval"? Our answer to this question is that any retrieval mechanism (and also any definition of "relevance") is never neutral, but is always considering some interests at the expense of other interests. To make a distinction between such views is to make a kind of classification, which is thus always necessary. To believe in a technical solution employing "relevance feedback" is a fallback to the positivist failure. The vision of automated feedback and value-free systems is tempting but based on problematic philosophical assumptions.

This ARIST chapter is based on the pragmatic understanding of concepts, meaning and semantics. This perspective may be able to address fundamental problems in KO and IR from a new and promising angle. The theoretical point of this paper is the view expressed by the American philosopher Hilary Putnam. He aptly gives a résumé of his criticism in a paper with the classical title “The meaning of ‘meaning’”:

> “Traditional semantic theory leaves out only two contributions to the determination of extension – the contribution of society and the contribution of the real world!” (Putnam, 1975).

Putnam is also known as a philosopher in the pragmatic tradition. We may thus list three characteristics by his (and our) philosophical point of departure:

- The focus on the relation between meaning and the real world (realism)
- The focus on the functional/pragmatic nature of meaning (pragmatism)
- The focus on the development of meaning in a social context (historicism and meaning collectivism/holism).

We can say with Putnam that these principles have been very much ignored in semantic theory and we can add that they have also been very much ignored in fields like Information Science that - as shown above - is very depending on semantics.

Semantics and subject knowledge

Advanced semantic tools demand proper subject knowledge for their design and administration as well as for their use and evaluation. This follows from the realist philosophical position formulated above: knowledge of semantic relations between terms requires world knowledge about the relations between the objects that the terms refer to. You cannot determine the semantic relations between the words “Copenhagen” and “Denmark” unless you know that Copenhagen is a part of Denmark.

This has been well known in research libraries and bibliographical databases as well as in education for librarianship. The Medline database, for example, demands: “A prospective indexer must have no less than a bachelor's degree in a biomedical science, and should also have a reading knowledge of one or more modern foreign languages. An increasing number of recent recruits hold advanced degrees in biomedical sciences.” (National library of Medicine, 2005).

Concerning the construction of ontologies for gene technology, Bada et al. (2004) writes:

> "One of the factors that account for GO's [Gene Ontology's] success is that it originated from within the biological community rather than being created and subsequently imposed by external knowledge engineers. Terms were created by those who had expertise in the domain, thus avoiding the huge effort that would have been required for a computer scientist to learn and organize large amounts of biological functional information. This also led to general acceptance of the terminology and its organization within the community. This is not to say that there have been no disagreements among biologists over the conceptualization, and there is of course a protocol for arriving at a consensus when there is such a disagreement. However, a model of a domain is more likely to conform to the shared view of a community if the modellers are within or at least consult to a large degree with members of that community".

The above quotations do not represent a new view. Already Richardson and Bliss considered the implications of the need of subject knowledge for education in librarianship and information science when they wrote:

> "Again from the standpoint of the higher education of librarians, the teaching of systems of classification . . . would be perhaps better conducted by including courses in the systematic encyclopedia and methodology of all the sciences, that is to say, outlines which try to summarize the most recent results in the relation to one another in which they are now studied together, ..." (Ernest Cushing Richardson, quoted from Bliss, 1935, p. 2).

M Lynne Murphy provides a linguistic investigation of semantic relations. Her conclusion is:

> "Plainly, the topic of lexical semantic paradigms has not been exhausted, and the metalinguisic approach discussed in this book gives rise to a number of new directions for lexicological research. It fits with (and exploits) a general trend in linguistic research to appreciate the particular relations that language engages in: the relation between language and context, language and conceptualization, language and linguistic behavior. While [Leonard] Bloomfield (1985[1936]) argued that linguists should ignore meaning because it is not properly "linguistic," to hold such a position in the current disciplinary context is untenable, since many if not most (if not all) linguistic phenomena cross boundaries between the linguist, the conceptual, and the communicative. In the case of lexical relations, this means that those who study it are not just linguists, but metalinguists." (Murphy, 2003, 242).

The domain analytic view in information science is an attempt to provide subject knowledge within the borders of LIS and in a way that still makes it possible to have a clear identity as information scientist (cf., Hjørland, 2002a). If librarians and information specialists are taught the content of a paper such as Ørom (2003), this should provide a better basis for all kind of information work related to the domain of Arts. In addition it should provide certain possibilities to generalize to other domains. In this way information specialists may provide knowledge, which is both domain specific and allow LIS to have a specific identity.

Domain knowledge is not only a problem for information science, but also for linguistics and many metasciences (such as cognitive science and the sociology of science). Much cognitive and linguist theory regarding concepts, meaning and semantics is strongly constrained by attempts to avoid "world knowledge".

That subject knowledge is important have theoretical implications for how concepts should be defined and semantic relations determined (whether by human or by machine). It has implications for an answer to the question: What kind of information is needed in order to determine the semantic relations between two terms A and B? This question will be considered in the next section.

Semantics and its "warrant"

Theories of semantics should be formulated in ways, which provides methodological implications for how to determine meanings and relations in semantic tools such as thesauri and semantic networks. Often such implications are not clear, which makes the theories vague and less fruitful. Murphy (2003, p. 111), for example, writes: "From the WordNet literature available, it is often difficult to determine the bases on which design decisions in WordNet are made. For example, Miller (1998b) notes that Chaffin et al. (1988) identified eight types of meronymy and Iris et al. (1988) distinguished four types, but he does not indicate how it was determined that WordNet should distinguish only three types". Similarly it is often unclear on what bases specific decisions are made in classification systems such as DDC or in thesauri such as "Thesaurus of Psychological Index Terms" (1st ed.: Kinkade, 1974; 8th ed.: Walker, 1997).

Frohmann (1983) is a paper about the semantic basis and theoretical principles of some classification systems. One of the important merits is that it is one of the rare papers showing that problems in classification should be seen as problems related to semantic theories. Concepts such as "dog", "cat", "whale", "pike" and "owl" may be grouped or classified in different ways. "For example, one principle of division divides the set according to nocturnal and diurnal characteristics. In this case, "cat" and "owl" belong to the first category, and the other terms to the second. Another principle of division separates mammals from non-mammals. In that case, "dog," "cat" and "whale" belong to the first category, whereas "pike" and "owl" belong to the second. Other divisions may be recognized (e.g., "land creatures," "water creatures" and "flying creatures")." (Frohmann, 1983, pp. 15-16).

Frohmann presents two semantic theories. The first one is that the categories to which a concept belongs are given *a priory* as part of the "meaning" of the term for that concept. The second one is that the categories to which a concept belongs must be found in the specific literature or discourse, of which the associating term is a part. Consequently the semantic relations are not given *a priory*, but are *a posteriori*. This distinction has implications for classification theory. Frohmann demonstrates that Austin's PRECIS system (as an example) is open to an argument from Wittgenstein's later philosophy of language. According to Frohmann the implication might be that systems of KO cannot be both machine-compatible and adequate, as Austin claimed (although he does not rule out other ways to construe systems that are both machine-compatible and adequate if based on other principles than those provided by Austin).

Thus a basic problem in KO is related to the problem of whether semantic relations are *a priory* or are *a posteriori:* whether they can be known before examining the literature or only after such an examination. What kind of literary warrant (or other kind of warrant) is needed in order to identify semantic relations and to classify concepts?

This question is also decisive for the question about the possibility of universal solutions to KO because *a posteriori* relations are unlikely to be universal. According to Frohmann *the Classification Research Group* (CRG) in England realized that semantic relations are *a posteriori* relations and have to be determined by examining the specific disciplinary literatures one after another. Neither Frohmann himself, nor the literature from CRG and Bliss Bibliographical Classification is, however, especially specific about precisely how concepts should actually be defined and their relations identified. Although it is correct that the CGR (and the Bliss

Classification System 2nd ed.) work on the basis of examining specific literatures, it is not clear (at least to this author) to what extent semantic relations are taken from the literature to be classified or are put down over that literature. My opinion is that those systems are based more on *a priory* principles than Frohmann suggests. There is a tendency within the whole facet-analytic tradition to work with universal categories like time and space and to classify the literature in relation to such pre-established categories. I believe this will be clearer when we analyses different theories of concepts and semantics below.

Let us look for some theoretical possibilities about the nature of concepts and semantic relations. They might be:

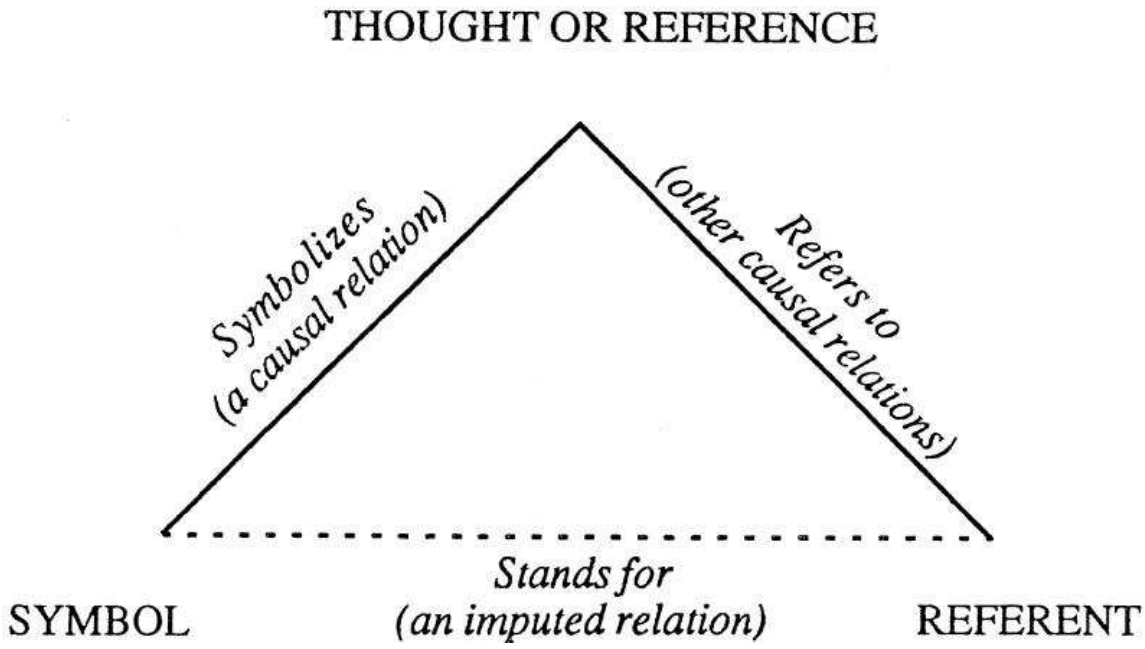
a) Query/situation specific or idiosyncratic
b) Universal, Platonic entities/relations
c) "Deep semantics" common to all languages (or inherent in cognitive structures)
d) Specific to specific empirical languages (e.g. Swedish)
e) Domain or discourse specific
f) Other (e.g. determined by a company or by a workgroup, "user oriented")

Before discussing these possibilities separately let us have some general considerations concerning the nature of semantic relations. Semantic relations are often displayed in standard lexica, for example, in *Longman Synonym Dictionary* (1986) and in WordNet and in similar semantic tools. It is well known, however, that, for example, synonyms are seldom synonyms in all contexts, but just in some contexts. It thus becomes important not to think of semantic relations as just "given", but to ask: "When are two concepts A and B to be considered synonyms? (or homonyms or otherwise semantically related). When is a semantic relation?". We should again ask the pragmatic question: What difference does it make whether or not, in a given situation, we choose to consider A and B semantically related in a specific way? This may look strange, given that many semantic relations seem intuitively "given" or authoritatively established in standard dictionaries. [3]

This relativity of meaning is also evident from Ogden & Richard's (1923)[4] famous triangle of meaning:

[3] Some texts (e.g., Foskett, 1977 and Dahllöf, 1999) define semantic relations as stable and different from "syntactic relations" (Foskett, 1977, p. 72) or from pragmatic relations (Dahllöf, 1999, p. 44). This is not in accordance with the theoretical view put forward in the present paper, and would make the question: "When is a semantic relation?" meaningless.

[4] Sowa (2000) writes about Ogden & Richards' (1923) triangle of meaning: "The triangle in Figure 1 has a long history. Aristotle distinguished objects, the words that refer to them, and the corresponding experiences in the *psychê*. Frege and Peirce adopted that three-way distinction from Aristotle and used it as the semantic foundation for their systems of logic. Frege's terms for the three vertices of the triangle were *Zeichen* (sign) for the symbol, *Sinn* (sense) for the concept, and *Bedeutung* (reference) for the object."



The Ogden and Richards (1923) semiotic triangle

The triangle implies that the referen~~ce~~ of an expression (a word or another sign or symbol) is relative to different language users: With the terminology of Peirce: "A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object* [or referent]. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen." (Peirce, 1931-1958, 2, 228).

Concerning a) Query/situation specific or idiosyncratic semantics

""When *I* use a word," Humpty Dumpty said,
in rather a scornful tone,
"it means just what I choose it to mean—
neither more nor less."
"The question is, "said Alice,
"whether you *can* make words mean
so many different things."
"The question is," said Humpty Dumpty
"which is to be master—that's all.""
Carroll (1899)

It is important to keep in mind that concept determination and semantic relations are to be used in, for example, query expansion (automatic or manual) as well as in query precision and query formulation. In a way it is the specific "information need" that determines which relations are

fruitful and which are not fruitful in a given search session. A semantic relation that increases recall and precision in a given search is relevant in that situation. Creative information searchers do just that: They provide search strategies that bring forward a fruitful set of identified documents by combining terms in unusual ways. Different terms may be combined using the Boolean operator OR in a given search. By implication are they regarded as equivalent terms (or synonyms) in the situation disregarding that they are normally not considered synonyms. For example are antonyms and contrasts different from synonyms. In information retrieval it is, however, often useful to search antonyms because given phenomena are often discussed in relation to their opposites. The implication is that in a given search it might be useful to regard antonyms as synonyms.

This pragmatic point of departure is important to keep in mind in developing a theory of concepts and semantics. Semantic relations are relative to a given task or situation and a given user of semantic relations may not share the same view of what terms are equivalent compared with people using the terms for other tasks. On the other hand is it clear that if we base a semantic theory on an individual/idiosyncratic view of concepts and semantics, then it is not possible to design systems for more than one user or situation, which is absurd. We need more stable principles on which to determine semantic relations. We need a semantic theory about the meaning of words in forms of typified practices. Knowledge about semantics in typified practices may then be used by information searchers in one way or another in order to include or exclude certain documents.

Concerning b) Universal, Platonic entities/relations

Mathematicians are, probably more than other professionals, relatively often Platonists. They believe that the mathematical concepts such as "π" (phi) have always existed and have just waited to be discovered. "π" is semantically related to the "radius" and the "perimeter" of a "circle" (because it is defined as the relation between those concepts). This semantic relation is universal and given (although the symbols chosen are conventional). According to Platonism, the meaningfulness of a general term is constituted by its connection with an abstract entity, the (possibly) infinite extension of which is determined independently of our classificatory practices. (cf., Haukioja, 2005)

The question for us is: Is it also *a priory* in the sense Frohmann (1983) meant? It may be sufficient to say that the semantics of, for example, mathematical concepts are not just given by intuition by the individual indexer. They have to be determined by considering the mathematical literature (or people educated in that literature). Even if the basic method in mathematics is a kind of rational intuition this does not imply that semantic relations in mathematic should be considered to be given *a priory* in knowledge organization.

Concerning c) "Deep semantics" common to all languages or inherent in cognitive structures (a priory relations).

Much research on semantics is based on the assumption that concepts are somehow "hardwired" to our mind or brain, for example, in our so-called "mental lexicon". This is perhaps most clearly seen in research on color concepts.

The book *Basic Color Terms: Their Universality and Evolution* (Berlin & Kay, 1969) has had a big impact on the view of color terms. In that book the authors claimed the universality and evolutionary development of eleven Basic Color Terms (BCTs): The following characteristics of this universalist view is written by one of the main critics of that view, Barbara Saunders (2000): "[T]he relation between Munsell, the workings of the visual system, and the colour naming behaviour of people, is so tight it can be taken to be a causative law. Diversity of colour-naming behavior is defined as a system-regulated stability evinced by Evolution. The full lexicalisation of

the human colour space is designated Evolutionary Stage Seven, as in American English; languages below this level are the fossil record".

Berlin & Kay's (1969) view of color concepts is contrasted with a cultural-relative view in which our color concepts (and semantics in general) are not supposed to be determined primarily by our visual system, but by our relative needs to act in relation to the coloured environment. "Sociohistorical psychology emphasizes the fact that sensory information is selected, interpreted, and organized by a social consciousness. Perception is thus not reducible to, or explainable by, sensory mechanisms, per se. Sapir, Whorf, Vygotsky, and Luria all maintained that sensory processes are subordinated to and subsumed within "higher" social psychological functions. "(Ratner, 1989)[5].

We may thus conclude that the universality of color terms is controversial. The dominant view is cognitivist and maintains the universality of concepts, while a well-argued minority maintains a relativist view of color concepts. This last view is related to the pragmatic view.

A certain version of "deep semantics" is the theory of semantic primitives according to which every word can be broken up into primitive kernels of meaning, *semantemes* (also called *semantic features* or *semantic components*). Semantemes are terms that are used to explain other terms or concepts, but cannot themselves be explained by other terms. The process of breaking words down into semantemes is known as *componential analysis* and has been most often used to analyze kinship terms across languages. The components are often given in more detail. For instance, kinship terms like those shown below might have three components: sex, generation, and lineage. Sex would be male or female; generation would be a number, with 0 = reference point's generation, -1 = previous generation, +1 = next generation; lineage would be either direct, colineal (as in siblings) or ablineal (as in uncles and aunts).

Word	**Semantemes**
Father	male + parent
Mother	female + parent
Son	male + offspring
Daughter	female + offspring
Brother	male + sibling
Sister	female + sibling

Cruse (2001, p. 8758) characterizes the theory of semantic primitives an "influential approach, much criticized but constantly reborn". He also writes (p. 8759): "In the earliest versions of componential analysis, the components were the meanings of words, and the aim of the analysis was to extract a basic vocabulary, in terms of which all non-basic meanings could be expressed. Generally speaking, the features recognized by earlier scholars had no pretensions to universality, and indeed were often avowedly language-specific. Later scholars aimed at uncovering universals of human cognition, a finite `alphabet of thought.' Accessible introductions to componential analysis can be found in Nida (1975) and Wierzbicka (1996)." According to Sparck Jones (1992, p.

[5] Regarding relativism in color concepts see in addition to Ratner, 1989 also Goodwin, 2000, Lucy, 1998, Roberson, Davies & Davidoff, 2000 and Saunders, 2000.

1609) was this theory influential in early thesaurus construction: "A thesaurus was seen as providing a set of domain-independent *semantic primitives*".

Theories about "innate ideas" (including concepts and semantic relations) have roots far back in the history of philosophy and are particularly connected to the rationalist philosophers (like Descartes and Leibniz). The theory of semantic primitives is also related to "logical atomism" (Oliver, 1998), versions of which were put forward by Wittgenstein, in his *Tractatus Logico-Philosophicus* (1921) and by Bertrand Russell (1924) both of which were affiliated with logical positivism. (As is well known Wittgenstein later changed his view towards a more holistic and pragmatic view of language). In linguistics Chomsky has been the main representative of this rationalist philosophy. The theory is similar to views put forward in LIS, e.g. in thesauri and in the facet-analytic tradition founded by Ranganathan as well as in "formal concept analysis" (cf., Priss, 2005).

Although this theory is dominating in the literature (and associated with the cognitive view), I do not find it plausible or fruitful for knowledge organization. First, the arguments that have been raised against it by the researchers mentioned above seem plausible. Secondly, the semantic relations in knowledge organization are mostly a product of scientific ontological models. The relations between chemical elements, for example, are not hardwired in our brains but are discovered by chemical researchers why the construction of KOS has to identify the semantic relations in the subject literature rather than in psychological studies.

Concerning d) Semantics specific to given empirical languages

A paper by Hedlund et al. (2001) is titled "Aspects of Swedish morphology and semantics from the perspective of mono- and cross-language information retrieval". The implication of what is said in the title is that the Swedish language has a semantic. In other words: Semantic relations are structural relations attributed to different empirical languages. This view is also evident in the literature about structural linguistics. As demonstrated in the figure below the English word "tree" has not the same meaning as the Danish word "træ". Given languages are structures in which the words classify the world differently.

<table>
<tr><th>English</th><th>*German</th><th>*Danish</th><th>*French</th><th>Italian</th><th>Spanish</th></tr>
<tr><td>Tree</td><td>Baum</td><td rowspan="2">Træ</td><td>arbre</td><td>albero</td><td>Árbol</td></tr>
<tr><td rowspan="2">Wood</td><td rowspan="2">Holz</td><td rowspan="3">bois</td><td rowspan="2">legno</td><td>Leña</td></tr>
<tr><td rowspan="3">skov</td><td>Madera</td></tr>
<tr><td>Woods</td><td rowspan="2">Wald</td><td>bosco</td><td>Bosque</td></tr>
<tr><td>Forest</td><td>forêt</td><td>foresta</td><td>Selva</td></tr>
</table>

Cultural relativity in word meanings

Originally presented by the Danish structural linguist Louis Hjelmslev, 1943.
Extended by information from Buckley (2001).

Also many techniques in computational linguistics and natural language processing (NLP) are based on structures that are specific for a given language. The commercial program Connexor is described in the following way: "It gives a semantic interpretation of the syntactic structure, which means that many language-specific patterns are normalized. For example, the Machinese representation of the sentence "A book was given to John" shows the notional roles object and indirect object that correspond to the similar roles in "Somebody gave John a book"". (Connexor, 2003-2004)

The focus of differences between different natural languages has been fruitful also for information science. Research such as Hedlund et al. (2001) has provided knowledge, which is very fruitful in relation to information retrieval. No doubt about that. On the other hand knowledge organizing systems (for example the UDC) are applied across many languages and developed field by field. Semantic structures may be established in different domains and may diffuse into general languages. Our conceptions of uranium and radium being radioactive materials are based on scientific discoveries made within physics and from here transferred into general language. Semantic structures in LIS cannot be established just by the study of natural languages because domain-specific knowledge is demanded.

Concerning e) Domain or discourse specific semantics

We have already described how pragmatism view descriptions and conceptions of objects as always made from certain perspectives and from certain pre-understandings and interests. This principle is also emphasized in, among other epistemologies, hermeneutics and Thomas Kuhn's theory of scientific paradigms. Although objects have objective properties the representations of those properties in languages and concepts are always more or less “subjective” or “biased” by individuals, social groups or by different cultures. Different human interests emphasize different properties of objects: pharmacology and chemistry, for example, emphasize different properties of the same chemical elements (a chemical database emphasizes structural descriptions while a pharmacological database emphasizes medical effects).

The implication is that semantic relations in given descriptions are reflecting some human interests: Pharmacology as a domain or discourse community emphasizes for example those semantic relations that are related to medical effects and side-effects. This does not imply that all semantic relations are domain-specific. Pharmacology as a domain is heavily depending on chemical research and the two domains share many concepts and semantic relations. Still, parts of their descriptions contain descriptions and semantic relation that is a reflection of the specific goals of their respective domains.

How are the basic semantic structures determined within a domain? Frank C. Keil outlines some important developments in theories about concepts and semantics:

> “The history of all natural sciences documents the discovery that certain entities that share immediate properties nonetheless belong to different kinds. Biology offers a great many examples, such as the discoveries that dolphins and whales are not fish but mammals, that the bat is not a kind of bird, that the glass “snake” is in fact a kind of lizard with only vestigial limbs beneath its skin. In the plant kingdom it has been found, for example, that some “vegetables” are really fruits and that some “leaves” are not really leaves. From the realm of minerals and elements have come the discoveries, among others, that mercury is a metal and that water is a compound.
>
> In almost all these cases the discoveries follow a similar course. Certain entities are initially classified as members of a kind because they share many salient properties with other bona fida members of that kind and because their membership is in accordance with current theories. This classification may be accepted for centuries until some new insight leads to a realization that the entities share other, more fundamentally important properties with a different kind not with their apparent kind.
>
> Sometimes it is discovered that although the fundamental properties of the entities are not those of their apparent kind, they do not seem to be those of any other familiar kind either. In such cases a new theoretical structure must develop that provides a meaningful system of classification.

> There are many profound questions about when a discovery will have a major impact on a scheme of classification, but certainly a major factor is whether that discovery is made in the context of a coherent causal theory in which the discovered properties are not only meaningful but central" (Keil, 1989, p. 159).

The quotation shows that concepts and semantic structures depend on our world-views and theories, including those determined by scientific discoveries. It is also supportive for a scientific realism, that science uncovers deeper and deeper layers of reality and in the process has to change our theories, concepts, classification schemes and semantics. Such a view is very different from prevailing views that concepts are inherent in the mind or in specific languages.

In the literature of any domain are different theories and epistemologies at play (cf., "Domains" in Hjørland & Nicolaisen, 2005). In some cases (e.g. in psychology) there exist different "schools" or "paradigms" side by side with their own journals (cf., Hjørland, 2002b). In most cases, however, such different epistemologies or "paradigms" are not self-conscious, and do not have formally established information sources and communication structures. In the case of medicine, the movement known as *evidence-based medicine* may be considered a "paradigm" but there are no self-conscious alternative paradigms in medicine, which challenges this view[6]. In such cases the existence of different "paradigms" have to be demonstrated by analyzing different methodologies and assumptions made in the field and studies of different "paradigms" (e.g. by using bibliometric methods) are much more difficult to perform. A working hypothesis is that different theories, background assumptions and "paradigms" are at play in <u>any</u> field of knowledge (although, of course, the degree of consensus varies from field to field, why different views may be almost absent in some fields).

The meanings of given words or symbols are mostly influenced by the dominant view or paradigm within a given domain or discourse. Any attempt to change the dominant view implies a need to reconsider the meanings. Often this is not clear to the users of those words and symbols: they may use terms and symbols with meanings that counteract what the users try to accomplish. When the need to redefine symbols has become clear to the users, they may choose to use a different term or to continue to use a term with a somewhat different meaning. In this way meanings are linked to different views, interests and goals, and terms should generally be considered polysemous[7]. Attempts to standardize terminology may unwittingly suppress certain views. This problem is, for example, important to consider in relation to *The Unified Medical Language System* (UMLS) project. Campbell et al. (1998) shows how ULMS have integrated the concept "Aspirin" from two different source-thesauri. They write:

> "It is obvious that the intension associated with a term in a source terminology is represented at least in part by its location in a hierarchy and by decisions made regarding synonyms and non-synonyms. Aspirin in the CRISP Thesaurus is a chemical; it is also a centrally acting drug that has antirheumatic, anti-inflammatory, analgesic, and antipyretic properties. Similarly, the UMLS equivalent of aspirin in SNOMED, acetylsalicylic acid, is a chemical. It is also a drug with several of the same properties that it has in the CRISP Thesaurus: It is a centrally acting agent, an analgesic, and an antipyretic. On the other hand, in SNOMED, acetylsalicylic acid is not synonymous with two other UMLS equivalents of aspirin, Easprin and Zorprin, because the first is a generic drug and the other two are proprietary drugs. Thus, in SNOMED, the intension of aspirin is clearly not the same as the intension of Easprin, yet aspirin and Easprin are linked to the same CUI. It may even be argued that there are subtle

[6] Perhaps "Narrative based medicine" (Greenhalgh & Hurwitz, 1998) should be considered a competing paradigm.
[7] This is clearly seen in the German tradition of "Begriffsgeschichte" to be introduced below.

> differences in the intension of aspirin in CRISP and SNOMED, yet these differences are obscured or lost when one moves from the source terminology to the CUI." (Campbell et al., 1998).

How a term like "aspirin" should be defined and which semantic relations should be assigned in a given KOS is thus not an objective fact, but a question related to the purpose of that KOS. As Campbell et al. (1998) write: "In that discussion we noted that most clinicians would probably not consider these three concepts [aspirin, Aspergum, and Ecotrin] interchangeable in the prescriptions they write. However, we also assert that from some possible perspectives, such as when we are concerned primarily with medication allergies, having these concepts all linked to the same extension makes perfect sense." In this way semantic decisions, such as whether aspirin, Aspergum, and Ecotrin should be considered synonymous terms have to be decided by considering the consequences such as whether these substances may substitute each other for the purpose that the KOS is designed to accomplish.

The implication of different “paradigms” for knowledge organization and semantics is that any bibliography of a certain size must confront conflicting ways of defining concepts and determine semantic relations. “Literary warrant” does not mean just to identify a text from which semantic relations may be inferred. The task is normally to negotiate between different claims put forward in different texts and to select the one, which have the highest degree of cognitive authority or is considered best in relation to the goal of the KOS. The information scientists producing a kind of KOS have to negotiate between different views more or less visible in the literature to be indexed. In practice this is often not done, however. The DDC, for example, claim to be based on the principle of literary warrant (Mitchell, 2001, p. 217). However, as Miksa writes: “It should also be mentioned that a kind of solution have been to arrange as many categories as possible in orders that reflected some kind of consensus among experts but thereafter simply doing something “practical” with the remainder. This appears to have been an approach characteristic of the DDC and the UDC as they developed over the years.”(Miksa, 1994, p. 149).

Systems like the DDC are conservative because it is uneconomic to make deep literary investigations and to change the system and in particular to reclassify books. Systems of this kind have to weight the advantages of an updated system in accordance with literary warrant on the one side and on the other side being a standard, which is only changed reluctantly. One may also say that such systems have to weight between being an optimal tool for the information seeker and a practical tool for the library manager. For the theory of information science it is nonetheless important to describe the principles of designing optimal search tools. And such principles have to deal with conflicting criteria of literary warrant.

Example: Should *social psychology* be classified with psychology or with sociology? Bibliometric arguments might claim that psychologists are dominating in social psychology, why it should be classified with psychology. Theoretical arguments might claim, however, that explanation of social psychological phenomena need to be founded in sociological theory, why is should be classified with sociology. Historical and bibliometric studies show that there are actually two social psychologies; *psychological social psychology* (mainly experimental) and *sociological social psychology*. Each of those social psychologies has its own courses, textbooks, journals etc., why a third possibility would be to distinguish between psychological and sociological social psychology. The point is that the kind of information presented here is necessary for any informed decision about how to classify. Exactly the same kind of information would be helpful for the information seeker (in order to discriminate between the two kinds of social psychology or in order to find related information). If a semantic tool should be optimized as a retrieval tool, such information

about conflicting views of semantic relations should be available. This would imply that classification research made such alternatives visible in the literature and that the construction of systems was based on such knowledge, with explicit references to and interpretation of literary warrant. One can say that the more is invested in designing classification systems, the more benefit for the user. Arbitrary solutions, easy solutions, standardized solutions or "practical solutions" from the administrative point of view does not provide the information seeker with insights in the structures of knowledge.

The implication of the existence of different "paradigms" is thus that any existing KOS can be examined in relation to what views have priority and what views are relatively repressed. As demonstrated by Ørom (2003) different KOS such as the UDC and the DDC are more or less biased towards different paradigms within (in this example) art studies. Although some systems (e.g. *the Arts and Architecture thesaurus*) are easier to adapt to new tendencies there are no neutral platforms or criteria on which to base classifications and semantic tools. Any semantic tool may be more or less in harmony or in conflict with the views represented in the literature. Which view should the designer choose? The majority view? (As with psychological social psychology). It is not possible to prescribe any view or any method for selecting one. If this was possible it would be possible to prescribe how to do science, which most philosophers of science find impossible. The only thing we can conclude is that a precondition of designing quality KOS is that the designer knows the different views and is able to provide a reasonably informed and negotiated solution. In addition the designer of KOS should provide pragmatic analyses of what goals the KOS is going to fulfill.

The information scientists should ask the pragmatic question: given the different interests and "paradigms" in the field, what kinds of interest should this specific system support? What difference does it make whether some kinds of semantic relations are used at the expense of others? The most important task of the information professional is probably to make the different interests and "paradigms" visible in the first hand in order to enable the user to make an informed choice.

f) Other kinds of warrant

In KO as well as in information science in general user-oriented and cognitive theories have flourished for some time. Do kinds of "user warrant" exist with regard to semantic relations? Beghtol (1986) discusses the following kinds of warrant:

- Literary warrant & terminological warrant
- Scientific/philosophical warrant
- Educational warrant
- Cultural warrant

She does not, however, discuss "user warrant" in this paper and it is also difficult to imagine that the establishing of relations between term A and B should be determined by investigating users (e.g. that the classification of whales as mammals should be determined by users rather than by experts). In the case of popular music (Abrahamsen, 2003) the experts on genre are generally not the musicologists because too few of them have specialized in this field. It is closer to the users' own expertise, but probably journalists are among the people determining and naming new genres (and thus meaning and semantics). Other kinds of warrant may exist. Albrechtsen & Mark Pejtersen (2003) claim a sort of *work domain* warrant. This view may represent a tendency to prefer oral sources to written sources in information science. Oral and written sources need the same kind of interpretation and argumentation. Information scientists may feel safer if they rely on "experts" compared to documents, but relevant documents are written by experts and are just as valid sources, if not more so.

Semantic relations

Semantic relations are the relations between concepts, meanings or senses. The concept [school] should be distinguished from the word 'school'. [School] is a kind of [educational institution]. This indicates a hyponymous or hierarchical relationship between two concepts or meanings, which is one kind among a long range of kinds of semantic relations.

The concept [School] may, for example, be expressed by the terms or expressions 'school' 'schoolhouse' and 'place for teaching'. The relation between 'school' and 'schoolhouse' is a (synonym) relation between two words, while the relation between 'school' and 'place for teaching' is a relation between a word and an expression. The relations between words are termed lexical relations[8]. 'School' also means [a group of people who share common characteristics of outlook, a school of thought]. This is a homonym relation: Two senses share the same word or expression: 'school'. Synonyms and homonyms are not relations between concepts, but are about concepts expressed with identical or with different signs.

Relations between concepts, senses or meanings should not be confused with relations between the terms, words, expressions or signs that are used to express the concepts. It is, however, common to mix both of these kinds of relations under the heading "semantic relations" (i.e., Cruse, 1986; Lyons, 1977; Malmkjær, 1995 & Murphy, 2003), why synonyms, homonyms etc. are considered under the label "semantic relations" also in this paper.

How many kinds of semantic relations exist? Is the number of semantic relations finite or infinite? What determines this number?

In the quotation below (Rosario & Hearst, 2001) it is stated that there are contradictory views in theoretical linguistics regarding the semantic properties of noun compounds (NCs). Some researchers argue that there exists a small set of semantic relationships that NCs may imply. Others argue that the semantics of NCs cannot be exhausted by any finite listing of relationships. Green (2001, p. 5-6) argues that the inventory of semantic relationships includes both a closed set of relationships (including mainly hierarchical and equivalence relationships) and an open set of relationships. Every time a new verb is coined, for example, the potential for the introduction of a new conceptual relationship arises.

Is it possible to make an exhaustive list of semantic relations? The answer is probably that any relation between objects (or processes or anything else) may be reflected in languages between the corresponding concepts. "Love" is a relation between some people, e.g. Tom and Clare. [Tom] and [Clare] are thus individual concepts with the semantic relation 'love'. [9] ('Tom' and 'Clare' are words which may refer to other individual concepts which do not share the same semantic relations). The limit to the number of semantic relations seems to be relations that nobody have found interesting to conceptualize. If this argument is correct then the number of semantic relations is infinite.

[8] "Lexical Semantics is about the meaning of words. Although obviously a central concern of linguistics, the semantic behaviour of words has been unduly neglected in the current literature, which has tended to emphasize sentential semantics and its relation to formal systems of logic". (Cruse, 1986).

[9] Such relations could be drawn, for example, in in semantic networks. See, for example, fig. 7 in McCann (1997).

Different domains probably develop new kinds of semantic relations continuously. Rosario & Hearst (2001) identified 38 semantic relations within medicine. [10]

> "In this work we aim for a representation that is intermediate in generality between standard case roles (such as Agent, Patient, Topic, Instrument), and the specificity required for information extraction. We have created a set of relations that are sufficiently general to cover a significant number of noun compounds, but that can be domain specific enough to be useful in analysis. We want to support relationships between entities that are shown to be important in cognitive linguistics, in particular we intend to support the kinds of inferences that arise from Talmy's force dynamics (Talmy, 1985). It has been shown that relations of this kind can be combined in order to determine the "directionality" of a sentence (e.g., whether or not a politician is in favor of, or opposed to, a proposal) (Hearst, 1990). In the medical domain this translates to, for example, mapping a sentence into a representation showing that a chemical removes an entity that is blocking the passage of a fluid through a channel. The problem remains of determining what the appropriate kinds of relations are. In theoretical linguistics, there are contradictory views regarding the semantic properties of noun compounds (NCs). Levi (1978) argues that there exists a small set of semantic relationships that NCs may imply. Downing (1977) argues that the semantics of NCs cannot be exhausted by any finite listing of relationships. Between these two extremes lies Warren's (1978) taxonomy of six major semantic relations organized into a hierarchical structure.
>
> We have identified the 38 relations shown in Table 1 [omitted here]. We tried to produce relations that correspond to the linguistic theories such as those of Levi and Warren, but in many cases these are inappropriate. Levi's classes are too general for our purposes; for example, she collapses the "location" and "time" relationships into one single class "In" and therefore field mouse and autumnal rain belong to the same class. Warren's classification schema is much more detailed, and there is some overlap between the top levels of Warren's hierarchy and our set of relations." (Rosario & Hearst, 2001).

Rosario & Hearst (2001) thus seem to support the view that the number of semantic relations is infinite.

Semantic relations resemble commonly used grammatical categories. Categories and grammatical relations represent abstractions. The former example "love" may thus be seen as a special case of "being affected" (An Aristotelian category). Although the number of semantic relations appears to be unlimited, in most cases a limited number of generalized kinds are used in practice.

In information retrieval the basic functions for semantic relations may be conceived as contributing to the increase of recall and precision. The inclusion of synonyms and broader terms in a query may, for example, contribute to increased recall. The differentiation of homonyms and the specification of terms may increase precision. In this way the widely use of standard semantic relations used in thesauri may be explained functionally. There are, however, recommendations that the number of relations should be expanded:

> "The participants [in a NISO 1999 workshop on standards for electronic thesauri] recommended that a much richer, hierarchically organized, set of relationships be developed. .

[10] Rosario & Hearst (2004) described the problems involved in distinguishing seven relation types between the entities "treatment" and "disease" in biomedical texts.

> . . There is reason to expect that provision for semantic relations in controlled vocabularies will become much more extensive in a future standard . . ." (Milstead, 2001, p. 65).

How should we explain this demand for a much richer set of relationships than ordinarily used in, for example, thesauri? The answer may imply a criticism of the traditional recall/precision way of understanding information retrieval. What information searchers need are maps that inform them about the world (and the literature about that world) in which they live and act. They need such maps in order to formulate questions in the first hand. In order to formulate queries and to interact with information sources are advanced semantic tools often very useful. This is probably especially so in the humanities, where concepts are more clearly associated with worldviews. In Germany the concept of conceptual history ("Begriffsgeschichte") is an illustration of this point: Historians and other humanist researchers have realized that in order to use sources from a given period, you have to know what the terms meant at the time. Therefore they have developed impressive historical dictionaries which provide detailed information about conceptual developments within different domains just as they have developed methodological principles on how to work with historical information sources (cf., Hampsher-Monk, Tilmans & Vree, 1998).

An example of a semantic tool developed in this tradition is *Reallexikon der deutschen Literaturwissenschaft* (Weimar, 1997-2003), which provide the following information for each term:

- The term (e.g. 'bibliographie')
- A definition (e.g. definition of 'bibliography')
- A history of the word (its etymology, e.g. the etymology of the word 'bibliographie')
- A history of the concept. (e.g. the history of the meanings of 'bibliography')
- A history of the field (e.g. the history of bibliographies themselves) and
- A history of research about the field (e.g. the history of research on bibliographies, i.e. library science)

This example is mentioned because it illustrates the existence of important work, which may inspire LIS to a broader approach to semantic relations. Not much research has investigated whether different domains need different kinds of semantic tools displaying different kinds of semantic relations. Roberts (1985) is an exception arguing for the importance for specific kinds of relations in the social sciences.

The 'intellectual' versus the social organization of knowledge

Are there semantic relations between citing papers and their cited papers? Some authors directly use this terminology (e.g. Harter; Nisonger & Weng, 1993; Qin, 1999, Song & Galardi, 2001). Other uses bibliometric methods in order to establish semantic relations in thesauri and information retrieval (e.g., Kessler, 1965, Pao, 1993; Rees-Potter, 1989, 1991; Salton, 1971 & Schneider, 2004), thus implying such a relation.

The way Harter, Nisonger & Weng (1993) examined semantic relations between citing and cited papers were by applying two methods: A macro analysis, based on a comparison of the Library of Congress class numbers assigned citing and cited documents, and a microanalysis, based on a comparison of descriptors assigned citing and cited documents by three indexing and abstracting journals, ERIC, LISA, and Library Literature. Both analyses suggest that the subject similarity among pairs of cited and citing documents is typically very small (at least in this domain). In interpreting this study it should be recalled that subject determination typically is a process with

great uncertainty and variance. If two documents, A and B, has a citing relation (directly or indirectly by co-citations or bibliographic coupling), they might be understood as semantically related whether or not they are assigned the same descriptors or classification codes by somebody (or whether or not they contains the same words, for that matter: one might, for example, be in English, the other in Danish). I will argue that the citing relation in itself is a kind of semantic relation. In order to do so, I'll make a distinction between 'intellectual' versus social semantic relations and argue that citing relations belong to the later.

The kinds of relations typically used in semantic tools are 'real' relations such as geographical relations (Denmark is part of Europe), biological relations (cats are mammals), and chemical relations (such as the relations implied by 'the periodical system'. Such relations are "ontological". Researchers produce ontological models and they are used to organize knowledge.

A "social relation" is a different kind of relation. For example, disciplinary relations are social. That sociology is classified as a social science means that sociologists belong to the community of social scientists. A discipline is a social concept defined as people with similar education or other social ties, such as sharing the same organizations and journals. Disciplines mostly have strong internal citation relations compared to relations to other disciplines. A citation network is thus a kind of social relationship.

In some cases ontological models of reality correspond very well with social organizations such as disciplines or citation networks. In other cases, the connections may be weak (many disciplines or "schools" may, for example, partly share ontological structures). Social constructivists tend to claim that ontological models and discoveries are just constructed, why the social organization of knowledge is somehow primary to the intellectual organization. Scientific realists, on the other hand, tend to see ontological structures as primary and the social structures as based on preexisting structures discovered by science.

Ontological models and theories developed by researchers as well as social organizations provide meaning to terms and semantic relations between terms. One may discuss which kind of meanings or relations are the most true or fruitful ones. However, as information scientists we provide semantic tools that are based on both kinds of relations. Bibliometric tools and tools based on ontological relations are available and in many cases supplement each other in information retrieval. We should study in what ways they supplement each other and not try to reduce one of them to the other. In other words: semantic relations as provided by citing relations are legitimate in their own right. They need not be verified the way that Harter; Nisonger & Weng (1993) and Schneider (2004) try to do it. A traditional thesaurus and a bibliometric map may, in different ways, inform a person seeking information. Their relative roles may be domain-dependent. A citation relation between too papers, A and B is in itself a semantic relation whether or not it corresponds with how A and B are otherwise determined to be related.

Conclusion

As formerly stated the pragmatic view of semantics suggests that words and expressions are tools for interaction and their meanings are their functions within the interaction, their aptitudes to serve it in their distinctive ways.

When information professionals classify documents or informational objects, the relevant meanings and properties are only available on the basis of some descriptions. This important consideration is emphasized by van Rijsbergen, 1979, and it is in opposition to an implicit assumption that prevails: that all relevant properties of the objects are obvious for the information specialists and that he or she follows some given principles, which provide an optimal classification, which is objective, neutral, and universal, just technical efficient. A textbook on classification such as Hunter (2002, p. 25) demonstrates how machine bolts may be classified

according to their material, thread size, head shape and finish. This example is probably not typical for classification of documents (it is classification made too simple). The same thing is often described differently for different purposes. Different human interests emphasize different properties of objects. A typical database, on which IR-experiments are performed, should be conceived as a merging of different descriptions serving different purposes.

Traditional approaches to knowledge organization have a greater affiliation with positivism than with the pragmatic view of semantics. The solutions provided have not been based on the view that a typical database, on which IR-experiments are performed, should be conceived as a merging of different descriptions serving different purpose and based on different epistemologies. The implication is that traditional views have provided solutions, which are at best statistical averages, which are not optimal to anybody. The possible prospect for knowledge organization based on a pragmatic understanding of semantics is a fine-tuning of KOS in different domains.

References

Abrahamsen, K. T. (2003). Indexing of Musical Genres. An Epistemological Perspective. *Knowledge Organization, 30*(3/4), 144-169.

Albrechtsen, H. & Mark Pejtersen, A. (2003). Cognitive work analysis and work centered design of classification schemes. *Knowledge Organization, 30*(3/4), 213-227.Antoniou, G. & van Harmelen, F. (2004) *A semantic web primer*. Cambridge, Mass: MIT Press, 1-149.

Bada, M., Stevens, R., Goble, C., Gil, Y., Ashburner, M., Blake, J. A., Cherry, J. M., Harris, M. & Lewis, S. (2004). A short study on the success of the Gene Ontology. *Journal of web semantics* 1(2), 235-240. Available at: http://www.websemanticsjournal.org/ps/pub/2004-9

Barnes, B. (2002). Thomas Kuhn and the problem of social order in science. IN: T. Nickles (ed) *Thomas Kuhn*, Cambridge University Press. (Pp. 122-141).

Bean, C. A. & Green, R. (Eds.). (2001). *Relationships in the organization of knowledg*e. Dordrecht: Kluwer Academic Publishers.

Beghtol, C. (1986). Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services,* 109-125.

Berlin, B. & Kay, P. (1969) *Basic Color Terms. Their Universality and Evolution,* Berkeley: University of California Press. (Reprinted 1991).

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American,* 284(5), 34-43.

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

Blair, D. C. (2003). Information retrieval and the philosophy of language. *Annual Review of Information Science and Technology*, 37. Medford (NJ): Information Today. 3 – 50.

Bliss, H. E. (1929).*The organization of knowledge and the system of the sciences*. With an introduction by John Dewey. New York: Henry Holt and Co.

Bliss, H. E. (1935). A system of bibliographical classification. New York: H. W. Wilson.

Bloomfield, L. (1936/1985). Language or ideas? IN: J. J. Katz (ed.): The philosophy of linguistics. Oxford: Oxford University Press (pp. 19-25). (Originally published in *Language, 12,* 1936)

Bloor, D. (1997). *Wittgenstein: Rules and Institutions*. London: Routledge.

Bonnevie, E. (2001). Dretske's semantic information theory and meta-theories in library and information science. *Journal of Documentation***,** 57(4), 519-534.

Brooks, T. A. (1995). "Topical Subject Expertise and the Semantic Distance Model of Relevance Assessment." Journal of Documentation, 51(4), 370-387.
Brooks, T. A. (1998). The Semantic Distance Model of Relevance Assessment. Proceedings of the 61st Annual Meeting of ASIS, Pittsburgh, PA, October 25-28, 1998: Information Access in the Global Information Economy, Vol. 35 (pp. 33-44) [pdf] [HTML]
Broughton, Vanda, Hansson, Joacim, Hjørland, Birger and López-Huertas, Maria J. (2005), "Knowledge organisation: Report of working group 7", in Kajberg, L. and Lørring L. (Eds), *European Curriculum Reflections on Education in Library and Information Science*, Royal School of Library and Information Science, Copenhagen, available at: http://www.db.dk/LIS-EU/workshop.asp
Buckley, G. (2001). Semantics. http://www.ling.upenn.edu/courses/Spring_2001/ling001/semantics.html (Visited July 31, 2005).
Budd, J. M. (2004). Relevance: Language, semantics, philosophy. *Library Trends, 52*(3), 447-462.
Campbell, K. E., Oliver, D. E., Spackman, K. A. & Shortliffe, E. H. (1998). Representing Thoughts, Words, and Things in the UMLS. Journal of the American Medical Informatics Association, 5(5), 421–431. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61323
Carroll, L. (1899). *Through the Looking Glass*. New York: M. F. Mansfield & A. Wessels.
Chaffin, R., Herrmann, D. J. & Winston, M. (1988). A Taxonomy of part-whole relations. Cognition and language, 3, 1-32.
Connexor (2003-2004). Machinese Semantics. http://www.connexor.com/software/semantics/
Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
Cruse, D. A. (2001). Lexical Semantics. IN: Smelser, N. J. & Baltes, P. B. (eds.) International Encyclopedia of the Social and Behavioral Sciences. Oxford. (Pp. 8758– 8764).
Dahlberg, I. (1978). A referent-oriented, analytical concept theory for INTERCONCEPT. *International Classification*, 5(3), 142-151.
Dahlberg, I. (1995). Conceptual Structures and Systematization. IFID Journal, 20(3), 9-24.
Dahllöf, M. (1999). *Språklig betydelse. En introduktion till semantik och pragmatik*. Lund: Studentlitteratur.
Daily, J.E. (1979). Semantics. In *Encyclopedia of Library and Information Science*, Volume 27, 209-215 (Reprinted in 2nd.ed., 2003).
Ding, C.H.Q. (2005). A probabilistic model for Latent Semantic Indexing. *Journal of the American Society for Information Science and Technology, 56*(6), 597-608.
Doerr, M. (2001). Semantic problems of thesaurus mapping. *Journal of Digital Information, 1*(8). URL: http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/
Downing, P. (1977). On the creation and use of english compound nouns. *Language*, 53, 810-842.
Dumais, S.T. (2004). Latent semantic analysis. *Annual Review of Inforemation Science and Technology*, *38*, 189-230
Eco, U. (1995). *The Search for the Perfect Language*. (Translated from Italian: Ricerca della Lingua Perfetta Nell a Cultura Europea). Oxford: Blackwell.
Ereshefsky, M. (2000). *The Poverty of the Linnaean Hierarchy : A Philosophical Study of Biological Taxonomy*. Cambridge: Cambridge University Press.
Fensel, D., Hendler, J. A., Lieberman, H., Wahlster, W. (Eds.). (2003). *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Cambridge: MIT Press.
Foskett, A. C. (1977). Assigned indexing I: Semantics. In: *The subject approach to information* (pp. 67-85). London: Clive Bingley.

Frohmann, B. P. (1983). An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG. *Cataloging & Classification Quarterly, 4*(1), 11-27. Green, R., Bean, C. A. & Myaeng, S. H. (Eds). (2002). *The semantics of relationships. An Interdisciplinary Perspective*. Dordrecht: Kluwer Academic Publishers.

Goodwin, C. (2000). Practices of Color Classification. Mind, Culture and Activity, 7(1-2), 19-36.

Greenhalgh, T. & Hurwitz, B. (1998). *Narrative Based Medicine: Dialogue and discourse in clinical practice*. London: Bmj Publishing Group.

Green, R. (2001). Relationships in the organization of knowledge: An overview. IN: Bean, C. A. & Green, R. (Eds.). Relationships in the organization of knowledge. Dordrecht: Kluwer Academic Publishers. (Pp. 3-18).

Green, R., Bean, C. A. & Myaeng, S. H. (Eds.). (2002). *The semantics of relationships. An interdisciplinary perspective*. Dordrecht: Kluwer Academic Publishers.

Hammerwohner, R. & Kuhlen, R. (1994). Semantic control of open hypertext systems by typed objects. *Journal of Information Science, 20*(3), 175-184.

Hampsher-Monk, I., Tilmans, K. & Vree, F. van (Eds.). (1998). History of Concepts: Comparative Perspectives. Amsterdam: Amsterdam University Press.

Harris, R. (2005). *The semantics of science*. London: Continuum International Publishing Group Ltd

Harter, S. P., Nisonger, T. E. & Weng, A. W. (1993). Semantic relations between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science, 44*(9), 543-552.

Haukioja, J. (2005). A middle position between meaning finitism and meaning Platonism. International *Journal of Philosophical Studies, 13*(1), 35-51.

Hearst, M. A. (1990). A hybrid approach to restricted text interpretation. In Paul S. Jacobs, editor, *Text-Based Intelligent Systems: Current Research in Text Analysis, Information Extraction, and Retrieval*, pages 38-43. GE Research & Development Center, TR 90CRD198.

Hedlund, T., Pirkola, A. & Kalervo, J. (2001). Aspects of Swedish morphology and semantics from the perspective of mono- and cross-language information retrieval. *Information Processing and Management, 37*, 147-161.

Hjelmslev, L. (1943) *Omkring sprogteoriens grundlæggelse*. København: B. Lunos bogtrykkeri a/s. (Many later editions and translations, e.g. *Prolegomena to a theory of language*. translated by Francis J. Whitfield. Baltimore : Waverly Press, 1953).Hjørland, B. (1998). Information retrieval, text composition, and semantics. Knowledge Organization, 25(1/2), 16-31. http://www.db.dk/bh/publikationer/Filer/ir_semant_2.pdf

Hjørland, B. (1997): Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science. Westport & London: Greenwood Press.

Hjørland, B. (1998). Information retrieval, text composition, and semantics. Knowledge Organization, 25(1/2), 16-31. http://www.db.dk/bh/publikationer/Filer/ir_semant_2.pdf

Hjørland, B. (2002a). Domain analysis in information science. Eleven approaches - traditional as well as innovative. *Journal of Documentation, 58*(4), 422-462. http://www.db.dk/bh/publikationer/Filer/JDOC_2002_Eleven_approaches.pdf

Hjørland, B. (2002b), Epistemology and the Socio-Cognitive Perspective in Information Science. *Journal of the American Society for Information Science and Technology, 53*(4), 257-270.

Hjørland, B. (2004b). Theory of knowledge organization and the feasibility of universal solutions. At: Eighth International ISKO Conference London July 16th 2004. http://www.db.dk/bh/Hjorland_Feasibility%20of%20universal%20solutions.ppt

Hjørland, B. & Nicolaisen, J. (2005). *The epistemological lifeboat*. Copenhagen: Royal School of Library and Information Science. Available: http://www.db.dk/jni/lifeboat/home.htm

Hjørland, B. & Nissen Pedersen, K. (2005). A substantive theory of classification for information retrieval. *Journal of Documentation, 61*(5), 582-597. http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Hjorland%20&%20Nissen.pdf

Hodge (2000). Systems of knowledge organization for Digital libraries. Beyond Traditional Authority Files. http://www.clir.org/pubs/reports/pub91/contents.html

Hunter, E. J. (2002). *Classification made simple. 2nd* ed. Aldershot: Ashgate.

Iris, M. A.; Litowitz, B. & Evens, M. (1988). Problems of the part-whole relation. IN: M. W. Evens (ed.): *Relational models of the lexicon*. Cambridge: Cambridge University Press (Pp. 261-288).

Iyer, H. (1992). Semantic interpretation of conjuncts - Boolean transformations. *International Classification, 19*(2), 72-76.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, Massachusetts: The MIT Press.

Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. American Documentation vol. 16, no. 3, 223-233.

Khoo, C. & Na, J.-C. (2005). Semantic relations in information science. *Annual Review of Information Science and Technology*, 40,

Kinkade, R. G (Ed.). (1974). *Thesaurus of psychological index terms. [1st.ed.]* Washington, DC: American Psychological Association.

Klaes, M. (2002). Some remarks on the place of psychological and social elements in a theory of custom. American Journal of Economics and Sociology, 61(2), 519-530. Available at: http://www.findarticles.com/p/articles/mi_m0254/is_2_61/ai_86469072#continue

Larsson, J. (2003). Finitism and symmetry. An inquiry into the basic notions of the strong programme. Göteborg: Göteborg University, Department of History of Ideas and Theory of Science. (PhD-dissertation).

Levi, J. (1978). *The Syntax and Semantics of Complex Nominals*. New York:Academic Press.

Longman Synonym Dictionary. (1986). Essex, UK: Longman.

Lyons, J. (1977). *Semantics. Vol. 1-2*. Cambridge: Cambridge University Press.

McCann, J. M. (1997). Generation of marketing insights: Semantic networks. Available at: http://web.archive.org/web/19990127092407/http://www.duke.edu/~mccann/mwb/15semnet.htm (Visited 2005-12-15).

Malmkjær, K. (1995). Semantics. IN: The Linguistics Encyclopedia. Ed. By Kirsten Malmkjær. London: Routledge. (Pp. 389-398).

Miksa, F. (1994). Classification. IN: *Encyclopedia of Library History.* Ed. by W. A. Wiegand & D. G. Davis. New York: Garland Publishing, Inc. (pp. 144- 153).

Miller, G. A. (1998). Nouns in WordNet. IN: C. Felbaum (ed.): *WordNet: an electronic lexical database*. Cambridge, NA: MIT Press (Pp. 23-46).

Milstead, J. L. (2001). Standards for relationships between subject indexing terms. IN: Bean, C. A. & Green, R. (Eds.). Relationships in the organization of knowledge. Dordrecht: Kluwer Academic Publishers. (Pp. 53-66).

Mitchell, J. S. (2001). Relationships in the Dewey Decimal Classification System. IN: Bean, C. A. & Green, R. (Eds.). (2001). *Relationships in the organization of knowledg*e. Dordrecht: Kluwer Academic Publishers. (Pp. 211-226).

Murphy, M. L. (2003). *Semantic relations and the lexicon: antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press. Book review1 http://www.sussex.ac.uk/linguistics/documents/studies_in_informatics.pdf book review2 http://www.sfs.nphil.uni-tuebingen.de/linguist/issues/15/15-686.html

National Library of Medicine (2005). Frequently asked questions. *Who are the indexers, and what are their qualifications?* http://www.nlm.nih.gov/bsd/indexfaq.html#qualifications

Nida, E. A. (1975). *Componential Analysis of Meaning: An Introduction to Semantic Structures*. The Hague: Mouton..

Ogden, C. K. and I. A. Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul.

Oliver, A. (1998). Logical atomism. IN: *Routledge Encyclopedia of Philosophy, Version 1.0*, London: Routledge.

Pao, M. L. (1993): Term and Citation Retrieval: A Field Study. Information Processing & Management, 29(1), 95-112.

Peirce, C. S. (1931-1958). *Collected Papers of C. S. Peirce* ed. by C. Hartshorne, P. Weiss, & A. Burks, 8 vols., Harvard University Press, Cambridge, MA.

Peregrin, J. (2004). Pragmatism and semantics. (Manuscript in English), published in German in: Fuhrmann, A. & Olsson E. J. (eds.): Pragmatisch denken, Ontos, Frankfurt a M., 2004, 89-108) http://jarda.peregrin.cz/mybibl/PDFTxt/482.pdf

Priss, U. (2005). Formal concept analysis in information science. *Annual Review of Information Science and Technology*, 40, pp.

Putnam, H., 1975, 'The meaning of "meaning".' In Language, Mind, and Knowledge*, K. Gunderson, ed. Minneapolis: University of Minnesota Press, pp. 131-193.

Qin, J. (1999). Discovering semantic patterns in bibliographically coupled documents**.** *Library Trends, 48*(1), 109-132.

Qin, J. (2000). Semantic similarities between a keyword database and a controlled vocabulary database: An investigation in the antibiotic resistance literature. *Journal of the American Society for Information Science, 51*(2), 166-180.

Ratner, C. (1989). A Sociohistorical Critique of Naturalistic Theories of Color Perception. *Journal of Mind and Behavior, 10*, 361-372. http://web.archive.org/web/20031029152929/http://www.humboldt1.com/~cr2/colors.htm

Read, C. S. (1973). General semantics. IN: *Encyclopedia of Library and Information Science*. Ed. By A. Kent, H. Lancour & J. E. Daily. New York: Marcel Dekker. (Vol. 9, pp. 211-221).

Rees-Potter, L. K. (1989). Dynamic thesaural systems: a bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management, 25*(6), 677-691.

Rees-Potter, L. K. (1991). Dynamic thesauri: the cognitive function. Tools for knowledge organisation and the human interface. Proceedings of the 1st International ISKO Conference, Darmstadt, 14-17 August 1990. Part 2, 1991, 145-150.

Roberson, D., Davies, I. & Davidoff, J. (2000) Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129*, 369-398.

Roberts, N. (1985). Concepts, structures and retrieval in the social sciences up to c. 1970. Social Science Information Studies, 5, 5567.

Rosario, B. & Hearst, M. (2001). Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy. P*roceedings of 2001 Conference on Empirical Methods in Natural Language Processing*, Pittsburgh, PA (EMNLP 2001). http://biotext.berkeley.edu/papers/emnlp01.pdf

Rosario, B. & Hearst, M. (2004). Classifying Semantic Relations in Bioscience Text . 42nd Annual Meeting of the Association for Computational Linguistics. Forum Convention Centre Barcelona. July 21-26, 2004. http://biotext.berkeley.edu/papers/acl04-relations.pdf

Russell, B. (1924) 'Logical Atomism', in Logic and Knowledge, ed. R.C. Marsh, London: Allen & Unwin, 1956, 323-43.

Salton, G. (1971). Automatic indexing using bibliographic citations. *Journal of Documentation, 27*(2), 98-110.

Saunders, B. (2000) Revisiting `Basic Color Terms.' *Journal of the Royal Anthropological Society, 6***,** 81-99.

Schneider, J. (2004) . *Verification of bibliometric methods' applicability for thesaurus construction.* PhD dissertation. Aalborg. Department of Information Studies, Royal School of Library and Information Science. Available at: http://biblis.db.dk/uhtbin/hyperion.exe/db.jessch04

Smith, B. (2004). Beyond concepts, or: Ontology as reality representation, *Formal Ontology and Information Systems* (FOIS), Amsterdam: IOS Press, 73–84.

Smith, B., Ceusters, W. & Temmerman, R. (2005). Wüsteria. *Proceedings of Medical Informatics Europe*. Available: http://ontology.buffalo.edu/medo/Wuesteria.pdf

Soergel, D. (2004). The Arts and Architecture Thesaurus (AAT). A critical appraisal. http://www.dsoergel.com/cv/B47_long.pdf

Song, M. & Galardi, P. (2001). Semantic relationships between highly cited articles and citing articles in information retrieval. ASIST 2001. Proceedings of the 64th ASIST Annual Meeting, 38, 171- 181.

Sowa, J. F. (2000). Ontology, Metadata, and Semiotics. Presented at ICCS'2000 in Darmstadt, Germany, on August 14, 2000. Published in B. Ganter & G. W. Mineau, eds., *Conceptual Structures: Logical, Linguistic, and Computational Issues*, Lecture Notes in AI #1867, Springer-Verlag, Berlin, 2000, pp. 55-81. Available at: *http://users.bestweb.net/~sowa/peirce/ontometa.htm*

Sparck Jones, K. (1970). Some thoughts on classification for retrieval. *Journal of Documentation, 26*(2), 89-101. http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Sparck%20Jones_1970.pdf

Sparck Jones, K. (1992). Thesaurus. IN: *Encyclopedia of Artificial Intelligence, Vol. 1-2*. Ed by S. C. Shapiro, New York: John Wiley & Sons. (Vol. 2, pp. 1605-1613).

Sparck Jones, K. (2005). Revisiting classification for retrieval. *Journal of Documentation, 61*(5), 598-601. [Reply to Hjørland & Nissen Pedersen, 2005] http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Sparck%20Jones_reply%20to%20Hjorland%20&%20Nissen.pdf

Stokolova, N. A. (1976). Syntactic tools and semantic power in information languages 2: Elements of a semantic theory of Information-Retrieval**.** *International Classification, 3*(2), 75-81.

Stokolova, N. A. (1977a). Elements of a semantic theory of Information-Retrieval 3.: Paradigmatic relations. *International Classification, 4*(1), 11-19.

Stokolova, N. A. (1977b). Elements of a semantic theory of Information-Retrieval. 1: Concepts of relevance and information language. Information Processing & management, 13(4), 277-234.

Talmy, L. (1985). Force dynamics in language and thought. In *Papers from the Parasession on Causatives and Agentivity*. Chicago: Chicago Linguistic Society. (21st Regional Meeting).

Thesaurus of ERIC Descriptors. 14th ed. (2001). Phoenix, AZ, Oryx Press.

van Rijsbergen, C.J. (1979), *Information Retrieval*, 2nd ed., Butterworths, London, also available at: http://www.dcs.gla.ac.uk/Keith/Chapter.3/Ch.3.html

van Rijsbergen, C. J. (1986). A new theoretical framework for information retrieval. In American Society for Computing Machinery, *Proceedings of the 1986 ACM Conference on research and development in information retrieval* (pp. 194-200). New York: ACM Press.

Vickery, B. C. & Vickery, A. (1987). Semantics and Retrieval. IN*: Information Science in Theory and Practice* (Chapter 6, pp. 133-179). London: Bowker-Saur.

Walker, A. (Ed.). (1997). *Thesaurus of psychological index terms. 8th*. ed. Washington, D.C.: American Psychological Association.

Warren, B. (1978). *Semantic Patterns of Noun-Noun Compounds.* Acta Universitatis Gothoburgensis.

Weber, M. (2005). How Strong is the Case for Social Relativism in Science? Lecture held at the Minnesota Center for Philosopht of Science. Available at: http://philosophy.duke.edu/pdf/weber_duke.pdf

Weimar, K. (Ed.). (1997-2003). Reallexikon der deutschen Literaturwissenschaft Band 1-3. 3. neubearb. Aufl. Berlin : Walter de Gruyter.

Wellisch, H. H. (2000). *Glossary of terminology in abstracting, classification, indexing, and thesaurus construction. 2nd.* ed. Medford : Information Today, Inc.

White, H. D. & McCain, K. W. (1998). Visualizing a discipline: An author co-citaiton analysis of information science, 1972-1995. *Journal of the American Society for Information Science 49*(4), s. 327-355.

Wierzbicka, A. (1996). *Semantics: Primes and Universals*. Oxford: Oxford University Press.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus.* London: Routledge and Kegan Paul Ltd. (English Translation of "Logisch-Philosophische Abhandlung," *Annalen der Naturphilosophie*, Ostwald, 1921). Hypertext of the Ogden bilingual edition: http://www.kfs.org/~jonathan/witt/tlph.html (Visited July 31, 2005).

WorldNet 2.1. A lexical database for the English language. Princeton: Princeton University, Cognitive Science Laboratory. Available at: http://wordnet.princeton.edu/

Ørom, A. (2003), "Knowledge Organization in the domain of Art Studies – History, Transition and Conceptual Changes", *Knowledge Organization*, Vol. 30 No. 3/4, pp. 128-143.

Appendix
Some important kinds of semantic relations which have been presented in the literature:

- Active relation: A semantic relation between two concepts, one of which expresses the performance of an operation or process affecting the other.
- Antonymy (A is the opposite of B; e.g. cold is the opposite of warm)
- Associative relation: A relation which is defined psychologically: that (some) people associate concepts (A is mentally associated with B by somebody). Often are associative relations just unspecified relations. In thesauri are antonyms, for example, usually not specified but may be listed, along with terms representing other kinds of relations, under "associative relations".
- Causal relation: A is the cause of B. For example: Scurvy is caused by lack of vitamin C.
- Homonym. Two concepts, A and B, are expressed by the same symbol. Example: Both a financial institution and an edge of a river are expressed by the word bank (the word has two senses).
- Hyponymous relationships (hyponym-hyperonym), also termed generic relation, genus-species relation or hierarchical subordinate relation. (A is kind of B; A is subordinate to B; A is narrower than B; B is broader than A).
- is-a (“instance”, example relation) designates the semantic relations between a general concept and individual instances of that concept. A is an example of B. Example: Copenhagen is an instance of the general concept 'capital'.
- Locative relation: A semantic relation in which a concept indicates a location of a thing designated by another concept. A is located in B; example: Minorities in Denmark.
- Meronymy, partitive relation (part-whole relation): a relationship between the whole and its parts (A is part of B) A meronym is the name of a constituent part of, the substance of, or a member of something. Meronymy is opposite to holonymy (B has A as part of itself). (A is narrower than B; B is broader than A).
- Passive relation: A semantic relation between two concepts, one of which is affected by or subjected to an operation or process expressed by the other.
- Paradigmatic relation. Wellisch (2000, p. 50): “A semantic relation between two concepts, that is considered to be either fixed by nature, self-evident, or established by convention. Examples: mother / child; fat /obesity; a state /its capital city”.
- Polysemy: A polysemous (or polysemantic) word is a word that has several sub-senses which are related with one another. (A1, A2 and A3 shares the same expression)
- Possessive: a relation between a possessor and what is possessed.
- Related term. A term that is semantically related to another term. In thesauri are related terms often coded RT and use for other kinds of semantic relations than synonymy (USE; UF), homonymy (separated by parenthetical qualifier) , generic relations and partitative relations (BT; NT). Related terms may, for example express antagonistic relations, active/passive relations, causal relations, locative relations, paradigmatic relations.
- Synonymy (A denotes the same as B; A is equivalent with B).
- Temporal relation: A semantic relation in which a concept indicates a time or period of an event designated by another concept. Example: Second World War, 1939-1945.
- Troponymy is defined in WordNet 2 in two senses: 1) the semantic relation of being a manner of does something 2) "the place names of a region or a language considered collectively".