

# ***Theory of knowledge organization and the feasibility of universal solutions***

***Birger Hjørland***

Eighth International ISKO Conference London

Friday July 16th 2004, Session 9B 2.15-3.45



# Overview of Presentation

- 1. Knowledge  
Organization and  
Concepts**
- 2. Knowledge  
Organization as  
Information Policy**
- 3. Groupings as  
Discoveries and  
Constructions**
- 4. Semantics and its  
"Warrant"**
- 5. Structural linguistics**
- 6. Compositional  
Semantics and  
Semantic Primitives**
- 7. Universal Solutions?**
- 8. Conclusion**



# 1. Knowledge Organization and Concepts

Classification and knowledge organization (KO) is about the grouping of like things.

In LIS “things” means physical documents or their parts or their representations as well as concepts used to characterize documents.

Any grouping of objects is based on conceptualizations of the things to be classified.



# 1. Knowledge Organization and Concepts

## Example 1:

Whales may be classified as “animals living in water” or as mammals. If whales are classified as mammals then documents about whales should be classified among other documents about mammals.

The distinction between classifying “things” and classifying documents and concepts is thus theoretically less important.



# 1. Knowledge Organization and Concepts

By implication KO is about the grouping of related concepts because our interaction with the world is always dependant on our pre-understandings and conceptualizations of the objects in the world.

People in different cultures and in different domains tend to conceptualize things differently, implying that they classify them differently.



# 1. Knowledge Organization and Concepts

From a sociocognitive point of view concepts are shared classifications internalized when people are socialized into a culture and learn a language or a subject. A concept is a way of classifying some part of the world (e.g. furniture).

Concepts are generally functional in facilitating the users interaction with the natural and cultural environment. Concepts serve pragmatic purposes.



# 1. Knowledge Organization and Concepts

From KO perspective the question is which concepts should be preferred in order to facilitate the goal of the information system.

The goal is decided in controlled vocabularies and classifications where a certain concept (or meaning) is selected while other concepts/-meanings/views are repressed relatively.

## 2. Knowledge Organization as Information Policy



It is the aim or purpose of the information system that determines what concepts/-meanings are most fruitful. In other words: this is in the end a political (or policy) decision.

No KO can ever be neutral! No KO can ever avoid bias!

To be neutral and without bias may be wrong ideals!

## 2. Knowledge Organization as Information Policy



Only in cases of universal consensus may neutrality and reduction of bias reflect sound ideals. Such cases are probably the exception rather than the rule.

As an alternative to neutrality and freedom from bias KO should be based upon an analysis of conflicting views and interests as well as upon explicit arguments on which basis the concepts should be defined and organized.

## 2. Knowledge Organization as Information Policy



### Example 2: The concept of hymn.

Recently the Danish Hymn book was revised. During the process it was discussed what to include. Some texts not referring directly to religious matters were also considered. In the end it is a political question what texts to classify and label “hymn”. Concepts facilitate certain human practices. Some people are conservative and some are progressive. Some represents a majority, some represents a minority.

## 2. Knowledge Organization as Information Policy



The most influential view tends to dominate the meaning. Often people representing minority views are not even aware that the concepts they use may oppose their own interests.

Nevertheless people use these concepts because they have learned to do so. If they try to develop a new practice (e.g. sing new texts in the church) they have to get support from other people and they have to change their concepts in order to facilitate the changed practice.

## 2. Knowledge Organization as Information Policy



### Example 3: Virgin Islands

In the Danish Dewey System (DK5) the Virgin Islands are classified as part of “Denmark” (46.5) even though they were sold to USA in 1916! This can be seen as a “bias” but it may be a well considered bias in that the users of the system will primarily be interested these islands because of their former relations to Denmark (and books are written on the former Danish possession).



### 3. Groupings as Discoveries and Constructions

Some groupings, e.g. whales are mammals not fish, are generally (but not universally) more fruitful than other groupings (A possible exception is books for children).

Many such fruitful groupings are discovered by science or they are otherwise constructed by human beings. They are not just 'given' (neither inborn in the human mind nor given by logical deduction), *but they are mostly documented in the literature.*



### 3. Groupings as Discoveries and Constructions

Precisely this fact gave name to our field of *Knowledge Organization* because its fathers, Cutter, Sayers & Richardsson, recommended that book classification followed scientific classification or what they termed “knowledge organization”.

Groupings and relations between concepts (i.e., semantic relations) are thus partly a result of scientific discoveries and the construction of theories and models.



## 4. Semantics and its “Warrant”

How then does LIS identify the criteria for classifying documents?

Two very fine papers about this problem, Frohmann (1983) and Beghtol (1986), have not, to my knowledge, attracted the attention they deserve – perhaps not even from the authors themselves. Both papers focus on the basis of the criteria used to identify what to consider “like” in KO.



## 4. Semantics and its “Warrant”

Frohmann (1983) is about the semantic basis and theoretical principles of some classification systems.

One of the important merits of this paper is that it shows that problems in classification should be seen as problems related to semantic theories.



## 4. Semantics and its “Warrant”

Concepts such as “dog”, “cat”, “whale”, “pike” and “owl” may be grouped or classified in different ways. Which way should be preferred in a given context?

One semantic view is that the categories to which a concept belongs are given *a priory* as part of the “meaning” of the term for that concept.



## 4. Semantics and its “Warrant”

An alternative view 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 priori*, but are *a posteriori*.

This is important for classification theory. Frohmann demonstrates that Austin’s system (for example) is open to an argument from Wittgenstein’s later philosophy of language.



## 4. Semantics and its “Warrant”

The implication might be that systems of KO cannot be both machine-compatible and adequate, as Austin claimed.

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.

This is also decisive for the question about the possibility of universal solutions because *a posteriori* relations are unlikely to be universal.



## 4. Semantics and its “Warrant”

Example 4: A library school could *a priori* be seen as a kind of school.

This seems to be “logical”, an universal true deduction. It is not something that must be decided by empirical analysis. It seems to follow from the meaning of the words “library school” and “school” respectively (or *at least* something that is generally true in the English language).



## 4. Semantics and its “Warrant”

Example 5: The PsycINFO thesaurus lists *neurolinguistic programming* (NLP) as a related term (RT) to *Neurolinguistics*. This is odd, because *a priori* it seems to be a narrower term (NT) rather than RT.

A study of the literature demonstrates that there are no co-citations: the literatures dealing with these two concepts are not related (Neurolinguistics is a part of “hard” science, NLP is a kind of alternative psychotherapy).



## 4. Semantics and its “Warrant”

Example 5 demonstrates that a priori semantic relations are often problematic.

They might, as this example demonstrates, be assigned by people lacking sufficient subject knowledge.

When semantic relations are decided in KO, those decisions must be based on some kind of information. This is sometimes called “*warrant*” (cf. Beghtol, 1986).



## 4. Semantics and its “Warrant”

Beghtol (1986) discusses the following kinds of warrant:

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

(Other kinds may exist. Albrechtsen & Mark Pejtersen (2003) claim a sort of *work domain* warrant).



## 4. Semantics and its “Warrant”

Here, the concept of “warrant” is used to identify the source of the semantic relations (implying that such relations are not self-evident or trivial). As Frohmann shows are such relations not just “given” (a priory) but develops in discourses and literatures. Consequently KO must have to study such discourses.

This is also indicated by Beghtol’s analysis of the concept of warrant and by example 5.



## 4. Semantics and its “Warrant”

As far as I know, nobody in LIS except Frohmann, have tried to connect grouping of like things to semantic theories about how meanings and relations are formed, and Frohmann mentions only one theory: that of the late Wittgenstein.

There is an obvious need to relate classification problems to semantic theories in general.



## 5. Structural linguistics

Hedlund et al. (2001) has as title: „*Aspects of Swedish morphology and semantics from the perspective of mono- and cross-language information retrieval*”

The title suggests that the Swedish language has a semantic of its own, i.e. one kind of semantic theory might suggest that semantic relations are inherent in specific natural languages.



## 5. Structural linguistics

If this theory is correct, then systems of KO should be made on the basis (“warrant”) on specific natural languages.

This is, however, seldom the case and the reason might be that this is a problematic theory of semantics: Semantic relations as used in KO are generally not inherent in natural languages, although structural linguistics sees languages as systems of meanings arbitrarily classifying the world in different ways.



## 6. Compositional semantics & semantic primitives

Another semantic theory views languages as consisting of semantic primitives, which may be combined to to complex meanings.

According to Sparck Jones (1992, p. 1609) this theory was influential in early thesaurus construction: “A thesaurus was seen as providing a set of domain-independent *semantic primitives*”.



## 6. Compositional semantics & semantic primitives. Example:

<b>Word</b>	<b>Semantemes</b>
Father	male + parent
Mother	female + parent
Son	male + offspring
Daughter	female + offspring
Brother	male + sibling
Sister	female + sibling



## 6. Compositional semantics & semantic primitives

This theory is also interesting in relation to KO because Ranganathan's view of subjects seems to be a version of this theory (although this issue has to my knowledge never been addressed!).

All examples used in the facet-analytic tradition treat subjects as composites of single "primitives" which can be freely combined (or combined according to a priori principles, e.g. sailing + Denmark + 1600).



## 6. Compositional semantics & semantic primitives

Henning (1995) wrote: “While componential analysis is useful for some exercises, it is not a representation of how language works; no linguist has ever been able to develop a complete list of semantic primitives . . .

If semantic primitives were to exist, they would number in the thousands and would resemble a mathematical logic system more than the mind's loom of language.”



## 7. Universal solutions?

Dominant trends in KO may thus have a problematic foundation which is related to much broader philosophical frameworks.

Umberto Eco (1995) wrote a book about the search for the perfect language which describes part of this problematic foundation. It is also related to rationalism and positivism in epistemology among other things.



## 7. Universal solutions?

It is an old rationalist dream to uncover the structure of the world as well as the structure of our knowledge *a priori*, once and for all. This dream is related to the search for perfect languages and perfect KO systems. Few people today regard this dream as based on solid grounds (Eco, 1995).



## 7. Universal solutions?

Principles for grouping and likeness as well as principles of concepts and semantic relations should not be based on positivist, *a priory* criteria. They should be based on qualified evidence primarily documented in specific subject literatures / discourses.

Such criteria are seldom universally accepted.



## 7. Universal solutions?

Universal systems of KO are feasible to the extent that the discourses they represents reflects universal conceptions. Clearly mathematical discourses are much more standardized and universal than, for example, religious or political discourses.

A system of KO may support or counteract various tendencies in the discourses on which it is based. It could, for example, have an standardization ideology of its own.



## 7. Universal solutions?

The negative side of such a universalizing ideology might be that minority views are repressed, that the development of new views are counteracted, and that users do not get detailed information about alternative conceptions and how to navigate between different views.



## 7. Universal solutions?

One possibility solution could be that different kinds of semantic tools should be developed to different fields.

Traditional standardized and universal tools like thesauri may be better suited for technological fields, while historical dictionaries in the tradition of *Begriffsgeschichte* may be better suited for the humanities.



## 8. Conclusion

The most important conclusion, as I see it is that different “paradigms” or conceptualizations in the discourses should be made visible and semantic tools should support the navigation not just between “topics” but also between “views” and interests. If such information is made explicit it might improve the possibilities for internationalization because nobody’s vital interests are then threatened.



## References

Albrechtsen, H. & Mark Pejtersen, A. (2003). Cognitive work analysis and work centered design of classification schemes. *Knowledge Organization*, 30(3/4),

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



## References

Eco, U. (1995). *The Search for the Perfect Language*. (Translated from Italian: *Ricerca della Lingua Perfetta Nell a Cultura Europea*); Oxford: Blackwell.

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.



## References

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.

Henning, J. (1995). Meaning. *Model Languages. The newsletter discussing newly imagined words for newly imagined worlds*. Volume I, Issue 6 (1/2) -- October/November 1, 1995. Available at: <http://www.langmaker.com/ml0106a.htm> (Visited 31 December, 2003)



## References

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).