UDC Seminar: The Hague: 4-5 June 2007

Preprint of the paper published in *Extensions & corrections to the UDC*, 29, 2007.

Progress in synthetic classification

Towards unique definition of concepts

Claudio Gnoli

(University of Pavia. Dep't of Mathematics. Library)

1: Helpful sequence

Classification systems are designed to contain all concepts into which a given domain of knowledge, or the whole of knowledge, is organized. Concepts are represented by both verbal terms, and their notational equivalents. The latter are devised in such a way that their ordinal sequence produces a helpful arrangement of the classified items [Ranganathan 1967, section EP].

Thus, the simplest classifications are basically lists of concepts arranged in a convenient order – for example, North-West to South-East:

A Iceland
B Norway
C Sweden
D Denmark
E Ireland
F United Kingdom
G Netherlands
H Belgium
I Luxembourg
J France
...

2: Hierarchy

When the items are many, however, it becomes impractical to browse all them: indeed, people find it difficult to scan lists longer than some tens of items, the average *futility point* being about 30 [Blair 1980]. This value is not far from the number of different letters in many alphabets, and maybe this is not chance. This makes alphabets very useful to represent classes. We can group some

items, and treat them as two-letter subclasses of a more general single-letter class:

```
A       America
E       Europe
EA          Iceland
EB          Norway
EC          Sweden
ED          Denmark
EE          Ireland
EF          United Kingdom
...
I       Africa
O       Asia
U       Oceania
Y       Antarctica
```

Some systems use digits instead of letters, but the underlying principle is just the same. To take an example from UDC:

```
(4)     Europe
(41)        British Isles
(410)           United Kingdom
(410.1)             England
(410.3)             Wales
(410.5)             Scotland
(410.521)               Orkney Islands
(410.522)               Shetland Islands
(410.523)               Western Isles
(410.7)             Northern Ireland
```

The combination of helpful sequence and hierarchy was about the totality of principles in which general classification schemes were based in the second half of the 19th century, when the first editions of the Dewey Decimal Classification (DDC) and the Library of Congress Classification (LCC) were published. These schemes we call *enumerative*, as they are basically an enumeration of all the concepts needed, including the compound ones, which are simply grouped into classes and subclasses, and listed in an helpful order. Enumeration can be considered the first step in the evolution of modern bibliographic classification [Ranganathan 1961; Sukiasyan 1998].

3: Common subdivisions

Although this can already seem enough to get a classification, soon it was seen that the enumeration of concepts produces many awkward repetitions. Indeed, several classes may be

divided according to the same principle, giving parallel subclasses:

EA	Norway
EBC	Norway **in the 17th century**
EBD	Norway **in the 18th century**
EBE	Norway in the 19th century
EBF	Norway in the 20th century
EI	France
EJL	France in the 15th century
EJM	France in the 16th century
EJN	France **in the 17th century**
EJO	France **in the 18th century**

This brings obviously to the idea of representing the concepts of centuries by a separate symbol, that can be reused to divide several classes in a consistent way. As we know, this device is widely used in DDC, where it was introduced in edition 2. A similar device had been already used half a century before in A.A.E. Schleiermacher's bibliographic classification, published in Germany in 1847 [Sukiasyan 1998].

In DDC, these recurring combinable concepts are called *common subdivisions*, thus emphasizing their function of subdividing classes into more specific subclasses. This is a top-down conception of classification, still far from the idea of building the whole classmark by synthesis of simpler concepts. Indeed, it is noticeable that in DDC we are allowed to append to a basic class not any number of common subdivisions, but only those judged sufficient to effectively subdivide the class. To index a serial dealing with rice cultivation, we are supposed to attach to the basic number for rice cultivation, `633.18`, the subdivision `-05` for serials, thus obtaining `633.180 5`. However, in case the serial only deals with rice cultivation in Italy, we will have to use the other subdivision `-0945` for Italy, thus obtaining `633.180 945`. The DDC manual forbids to attach both `-0945` and `-05`: only one subdivision is expected, and the rule says that place must prevail on bibliographical form. This means that a user searching for all serials dealing with agriculture by combining notation `63-` and `-05-` will fail to find this item. Maybe this is one reason why Ranganathan [1967, CX2] found that DDC did not produce coextensive classmarks.

In other schemes, like UDC and the Bliss Bibliographic Classification 2nd edition (BC2), very similar devices are called *common auxiliaries* – a term emphasizing more their function of being optionally appended to any basic class, than that of subdividing them. Indeed, there are fewer restrictions on combining several of them. The basic mechanism remains the same anyway.

Common subdivisions/auxiliaries are, in historical terms, the first way to produce *synthetic* classmarks, that is, to build classmarks by synthesis of simpler concepts, which may or may not occur in combination with the basic class. Classifications using this kind of system as a normal

means of combining concepts are called *analytico-synthetic*, to mean that the subject of a document is first analyzed into simpler separated concepts, then each concept is represented by the appropriate notation, and finally all are re-assembled to give the synthesized classmark. Any classification can be synthetic at a greater or lesser degree. Otlet and LaFontaine conceived UDC to be synthetic from its very beginning, by adding to the original DDC notation some auxiliaries and connecting devices; probably this feature of UDC influenced Ranganathan's later development of analytico-synthetic techniques [McIlwaine 1997]. Since the 1970s, DDC has also started to introduce similar synthetic mechanisms.

4: Facets

Many analytico-synthetic schemes are also described as *faceted.* This term was first introduced in library and information science in the 1930s by Ranganathan, who based his Colon Classification (CC) on facet analysis. His method was then adopted, spreaded, applied, and discussed by the Classification Research Group [1955], especially through building schemes for special domains [Vickery 1960].

In faceted classifications, each discipline or domain is analyzed into a certain number of facets typically occuring in it, and concepts are distributed and listed under the appropriate facets. The scheme then gives rules to combine facets, including those for citation order. In this way, synthesis is no longer an auxiliary tool, but the main and usual way of obtaining classmarks. The schedules get even shorter, as they are less and less enumerative, the concepts occurring in a discipline being listed only once under their own facet.

Though not being completely faceted yet, UDC provides facets in some classes and auxiliaries:

`(410.5)`	Scotland
`(410.5-15)`	rural Scotland
`(410.5-22)`	western Scotland
`(410.5-15-22)`	rural western Scotland

5: Phases

Although making a substantial progress in synthesis, facets do not provide for any possible combination of concepts. Indeed, they only work within the main class (often corresponding to a discipline) for which they are defined, allowing the combination of only the concepts occurring in it. On the other hand, subjects sometimes include a combination of concepts from different

domains. The class of animal communication may have facets such as sender, addressee, channel, code, function, meaning, and environment, by which we can build a faceted classmark for compound concepts like "courtship songs of passerines in forest habitats". But when our document also deals with the application of acoustic analysis to the study of courtship songs, then we need to combine the concepts from biology with others from physics and sound technology.

This kind of combination has been called *phase relation*, where the phases are the single components (each being faceted or not) coming from different domains. Ranganathan [1967, PE] acknowledged five kinds of phase relation: general, bias, comparison, difference, and influencing, though admitting that "other kinds of phase relations may be recognized in course of time". Perreault [1965] reviewed of many kinds of relations used in indexing systems.

In UDC, phase relations can be expressed by connecting symbols like the colon and plus sign:

`338.48`	tourism
`599.5`	marine mammals
`338.48:599.5`	tourism related to marine mammals

The colon exactly means a reciprocal influence; more specific phase relations, similar to those listed by Ranganathan, can be expressed by subdivisions of the common auxiliary `-042`:

`(410.5-15-22)`	rural western Scotland
`502.172`	wildlife conservation

`338.48:599.5(410.5-15-22)-042.3:502.172`
tourism related with marine mammals in rural western Scotland influenced by wildlife conservation

This compound subject is that of a real paper, entitled "The value of conserving whales: the impact of cetacean-related tourism on the economy of rural West Scotland" [Parsons et al. 2003]. It was encountered during the ILC tests described below, and appears to be a suitable example for a discussion of complex compounds.

UDC-like phase relations have also been incorporated in FATKS, a recent project of an analytico-synthetic classification scheme especially focused on the humanities [Broughton & Slavić 2007]. In FATKS, as notation is not conditioned by the historical heritage like in UDC, several special symbols can be defined to express phase relations, resulting in even more expressive classmarks:

`590+420`	education in addition to religion
`420/590`	the field of subjects spanning from education to religion
`590:420`	religion and education in coordinate (reciprocal) relationships
`590=420`	comparison between religion and education
`590»420`	the influence of religion on education
`420«590`	religion for educational purposes

590-420 religion as viewed by education
590<420 education as part of religion

Expressive notation, as in UDC and FATKS, is especially useful for information retrieval in a digital environment. Indeed, the system can search for any classmark containing e.g. the string 502.172 in any position, and display all subjects connected in some way with wildlife conservation, independently from the context in which they appear [Slavić 2006]. This is much more difficult in analytico-synthetic classifications with a non-expressive notation, like BC2 [Slavić & Cordeiro 2004].

6: Subject device

CC also uses a *subject device*, by which any class can act as a specifier of another class. To do this, the former is enclosed in brackets:

C physics
G biology
G:(C) biophysics

O-,2J64,51:*g* criticism of *Hamlet*
S psychology
O-,2J64,51:*g*(S) psychology of *Hamlet*

"CC uses Subject Device quite often. For example, it is used in the individualization of: (1) some substances in Organic Chemistry; (2) some buildings in Architecture; (3) some subjects in Sculpture; (4) special views in Metaphysics; (5) subjects in teaching techniques; and (6) industries in Economics" [Ranganathan 1967, ND].

In some classes of DDC and UDC, "add to" and "divide like" instructions are found, by which a class is subdivided according to the subdivisions of another class. This, however, does not hold as a general rule valid throughout the whole scheme. A more systematic device is UDC table -029 of "properties derived from other main classes", where the digits following -029: are taken from any class in the scheme:

-029:001 scientific properties
-029:1 philosophical properties
-029:133 occult properties
-029:51 mathematical properties
-029:57 biological properties

This system is also reproduced in FATKS notation (M9):

110	science of science
(M9:110)	scientific properties
120	philosophy
(M9:120)	philosophical properties

7: Place of unique definition

Phase relations, the subject device, and similar mechanisms allow the combination of any class with any other, and thus go very much in the direction of a fully synthetic classification. Still, there are some cases where compound concepts are not expressed in this way, but a concept has a different notation in different parts of the scheme. Let's consider an example in UDC. Among the common auxiliaries there is a table for materials, including -034.1 “ferrous metals”. Thus,

73	plastic arts
-03	common auxiliaries of materials
-034.1	ferrous metals, iron and steel
73-034.1	plastic arts using ferrous metals

If one searches for ferrous metals as materials, one can retrieve all items where they appear in combination with any concept, e.g. ferrous sculptures but also ferrous engines. However, this notation is different from those marking ferrous metals and iron in other parts of the scheme:

-032	naturally occurring mineral materials
-032.41	iron and manganese ores
546	inorganic chemistry
546.72	iron
553	economic geology, mineral deposits
553.3	ore deposits (metalliferous mineral deposits) in general, iron and manganese ores
669	metallurgy
669.1	ferrous metallurgy
67	various industries, trades and crafts
672	articles of iron and steel in general
68	industries, crafts and trades for finished or assembles articles
682	smithery, blacksmithery, farriery, hand-forged ironwork

The searcher is supposed to want information about ferrous metals in only one context of the

possible ones of materials, inorganic chemistry, economic geology, metallurgy, or crafts. She cannot search for just ferrous metals as such, combined with any other class, because the notation to be searched is different in each case. This reflects the assumption that users always have a preferred disciplinary perspective: they are doing a research either in art, or in chemistry, etc. Although this is often the case, this situation discourages interdisciplinary approaches and discovery of new relations between disciplines not canonically related.

This assumption by classification designers contrasts with the wishes, coming from many parts [Beghtol 1998; ISKO Italia 2007; Szostak 2007], that knowledge organization serves interdisciplinary exploration and discovery. To overcome the limitations produced by the disciplinary boundaries, a possible way is to define classes as phenomena, instead of objects of study as is done in our usual disciplinary classifications. Such a radical innovation was attempted by the Classification Research Group in a research project granted by NATO in the 1960s. Austin [1969, p. 151] explained some reasons why disciplinary classes are problematic: “A new idea or a new discovery may not yet have appeared in that part of the schedules we happen to need, even though it appears somewhere else. For example, although 'iron' may appear in the chemistry tables, we might have to wait while the editor establishes the right number for iron in, say, a schedule of drugs”.

In order to classify phenomena independently from disciplines, the CRG arranged them according to the integrative levels to which they belong: atoms, molecules, cells, organisms, societies, etc. This gives a single reference place where a phenomenon appear in the schedules: iron is primarily a chemical element, therefore its notation will be defined at the integrative level of atoms, though it will also appear in compounds in mining, metallurgy, crafts, arts, etc. “The arrangement of entities should ideally be such that each entity has, as its primary place, the one where its relations with neighbouring entities are constant – in the term of J. E. Farradane, its place of 'unique definition', where all characteristics essential to the definition, and no more, are available. Fewer would be inadequate, more would be superfluous. For example, a horse is only sometimes a sporting animal or draft animal, but is always a perissodactyl mammal” [Foskett 1970, p. 25; see also Tomlinson 1969, p. 29].

Jason Farradane, who is credited with the notion of place of unique definition, was also a CRG member. He had published radically innovative papers proposing a “relational indexing”, where classmarks were synthesized by a set of operators connecting the notation of each concept: “Each isolate will have a constant symbol in whatever aspect or context it is used, and no symbol will represent more than one isolate” [Farradane 1952, p. 90]. Although Farradane's system has not spread extensively, his notion of unique definition has great theoretical importance in the perspective of synthetic classification, or, in his terms, “inductive notation”. If iron were always

expressed by the same string in all contexts, it would be possible to find all its occurrences independently from disciplinary boundaries. Users interested only in iron as an ore could still impose limits on the search results, by combining the notation for iron with that for mining, but would not miss results on e.g. the economy of iron ores.

Research on classification by phenomena has been resumed by the Integrative Level Classification (ILC) research project, which is testing this techique by applying it to bibliographic databases in various domains [Gnoli & Hong 2006]. The most recent and extended example is the BioAcoustic Reference Database, maintained by the Interdisciplinary Centre of Bioacoustics and Environmental Research in Pavia, which includes more than 3,000 papers dealing with animal communication and especially cetacean studies. We already mentioned one paper from this archive, "The value of conserving whales: the impact of cetacean-related tourism on the economy of rural West Scotland", as it represents a series of matters in synthetic classification well.

As we have seen, UDC can cope well with such complex compounds, thanks to its synthetic features. However, UDC classes are defined according to disciplines: `338.48` tourism is a subclass of `33` economics, although tourism is also related to other disciplines, like entertainment and geography, and touristic concepts like free camping seem to have little to do with economics. A disciplinary scheme is forced to make a choice between disciplines, while a phenomenon scheme defines a phenomenon independently of any particular discipline: not tourism economy or tourist entertainment or tourist geography, but simply tourism.

It can be noticed here that even DDC, though explicitly listing the same concept under different disciplines with different notations, gives priority to one of them in the alphabetical index: that one, indeed, is listed before all the disciplinary occurrences of the concept, as the "interdisciplinary number" for that concept, that is, its implicit place of unique definition. This number is that chosen by the Italian Central National Library of Florence to be recorded as the DDC reference number of each term, in the thesaurus of the Nuovo Soggettario [Lucarelli *pers. comm.*]. Thus, although notations for horses appear in zoology, husbandry, sport and transport history, the interdisciplinary number for horses is that under zoology. Still, as DDC is a disciplinary scheme, it is possible to express literally "horse zoology", but not just "horses".

8: Relations between common auxiliaries and main classes

In UDC, the concept of Scotland is taken from the table of common auxiliaries of place. This table is formally autonomous, but practically it is connected with the class of geography: indeed, to index a book on Scotland, one is supposed to combine notation `913` for regional geography with

auxiliary (410.5) for Scotland; after a recent revision, general history is also divided by place auxiliaries. In a similar way, in linguistics, the studies on a particular language are expressed by attaching to the basic notation 811 the numbers taken from the table of common auxiliaries of language. Therefore, common auxiliaries and main classes are variously related.

"One technical question had to be resolved: what was the appropriate location for a particular array? A reasonable solution was that the most exhaustive and highly ordered lists of recurring concepts should be positioned in the main class where they were most commonly used. Let us pose the questions our colleagues asked in the past, find the answers and see how simple and rational their solutions were. Where should the exaustive list of languages appear? – Obviously, in linguistics. Where should we list place, countries and continents? – Evidently, in geography. Where should historical periods or chronological concepts be classified? – No doubt, in history. All that was needed was a hint on how to synthesize the appropriate characteristics in a single class number" [Sukiasyan 1998].

Links between common auxiliaries and main classes are made explicit and formalized in the database of FATKS: "Wherever relevant, concepts are linked across all three sections of the classification vocabulary with 'see also' references, thus creating a semantic network that links the whole of knowledge with humanities, and both of these with the common auxiliaries. [... U]nlike traditional faceted classifications, the FAT-HUM system is built on the principle that each concept must have a permanent data representation (i.e. notation) no matter to which syntax combination it engages" [Broughton & Slavić 2007].

In the perspective of unique definition, I suggest that these links could be further improved by reversing their definition: rather than defining the subclasses of geography in terms of the common auxiliaries of place, one could define common auxiliaries of place in terms of the main classes for regions. In the ILC draft scheme, no independent tables of common auxiliaries are planned: indeed, even concepts to be used frequently in combination with others, like places, have to be taken from their place of unique definition in the main table. Territories and places are defined within the class K99 "places of the contemporary Earth", and all facets of place anywhere in the scheme take their notation from there. To express tourism in Scotland, one can attach to Xy "tourism" the facet 2eec, taking its notation from the list of regions within main class K:

K99eec	Scotland
2 [K99]	common facet of place
Xy2eec	tourism in Scotland

By the same mechanism, one gets

`Nyw`	rural environments
`25 [Ny]`	common facet of environment
`Xy2eec25w`	tourism in rural Scotland
`Mqvtn`	cetaceans
`S3w`	wildlife conservation
`S4qvtn3w`	conservation of cetaceans
`S4qvtn3w)86(Xy2eec(4s)25w)`	conservation of cetaceans affecting tourism in rural western Scotland

Notation in square brackets, called *extra-defined foci*, indicates where the value of facets has to be taken from; as it is omitted in the compound classmarks, these result in shorter notations [Gnoli 2006]. Script instructions have been included in the search interface so that, when searching for `Nyw` "rural environments", the classmarks containing the fragment `25w` be also included in the results [Gnoli & Hong 2006].

9: Dependence

While it is extremely useful to record links between related concepts, and express subjects by the analytico-synthetic method, this has its limitations too. Indeed, if each concept were analyzed into its most elementary elements (provided they were identifiable), it could miss its peculiar meaning. Sculptures can be made of iron, and iron atoms are made of quarks, but the user searching for iron sculptures clearly will not be interested in information on their quarks. The theory of integrative levels itself claims that each level has its own relevant properties, which cannot be expressed in terms of the lower constituting levels: elasticity is a property of iron artifacts, but not of quarks. Thus, classificationists and classifiers must find each time the appropriate level of analysis, to know which relations are worth to be represented in notation, which others should be just recorded as links, and which ones should be omitted. It is probably appropriate that, in the UDC time auxiliaries, "`638`" "iron age" has a notation independent from that of iron ores. Iron ores and iron age are ontologically connected, but the two concepts are significantly different. This kind of connection can be better recorded in the schedules as an associative relation, e.g. "iron: *see also* iron age". From the perspective of integrative levels, it is a *dependence relationship* [Gnoli et al. 2007], that is, the existence of the iron age depends on the existence of iron ores, although other factors (like human populations and toolmaking) contribute to it, and it lies at the higher integrative level of human history.

Another case is the relation between languages and their countries of origin. English language `811.111` originated in England `(410.1)`; however, it could be inappropriate that it be

represented by the same fragment of notation (i.e., English = England + language). This is because languages and countries do not correspond exactly (English is spoken not only in England; Basque is spoken in both Spain and France), and the helpful sequence of their two arrays can be different (for its linguistic properties, Basque should not be listed near Indoeuropean languages); in the same way, in CC, the array of cultivated plants in agriculture is different from that of plants in botany, as their most relevant characteristics are not the same.

Nevertheless, it can be useful for some dependence relations to be recorded in the schedules, and they can therefore be displayed and navigated or used for query expansion when requested. Examples of this in ILC would be:

`K99eey`	England	
`Quyemi`	English language	« `K99eey` England
`K99oa`	Australia	
`Quys`	Australian aboriginal languages	« `K99oa` Australia

The user searching for Australian aboriginal languages will thus get only the items dealing with the languages; however, she will get a hint that these languages are connected with the territory of Australia, so that she could decide whether to expand her search and include items about Australia, especially if the first search has retrieved few items: indeed, a general work on Australia could include information and references to Australian languages. All these possibilities imply the notion of unique definition.

10: Attributes

The CRG draft scheme of phenomena actually consists of two tables: one for entities, arranged in order of integrative levels, and the other for *attributes*. The latter includes *general relative terms*, like "very" or "superior", *positional terms* of time, space, and person, *properties* like mass, colours and flavours, and *activities* such as resistance, origin, motion, etc. [Austin 1969, p. 148-160]. These have to be combined with the notation for entities, like in "cetaceans : conservation". Such partition of phenomena has been preserved in the BC2 classes for phenomena treated in an interdiscipinary way, which are divided into `4` "attributes", `5` "activities and processes", and `6` "entities" [Gnoli 2005].

Although attributes are not defined as classical common auxiliaries, they behave in a similar way. We are thus back to a scheme composed of several tables, each containing concepts not connected with the other tables. This seems inconsistent with the logic of integrative levels, as each

attribute also belongs to a given integrative level: “colour” is defined at the level of energy radiations, “gender” at the level of organisms, “shape” at the level of geometrical dimensions, etc. The property of gender can be attached to many different things, but not to things at levels lower than organisms, as it emerges at that level: you can have female cetaceans or female doctors or female musicians, but not female lakes.

If a scheme is based on a list of integrative levels, it would seem a further step if properties and processes were defined in the schedules together with the entities of the appropriate level (and then reused in the higher levels where needed). They would share with the entities the first part of their notation, while further digits would express the fact that they are properties, or activities, or processes:

```
M       organisms
M9          organisms of gender
M9f             female
M9m             male
```

In recent UDC updates, common auxiliaries have been added to express properties `-02`, relations, processes, and operations `-04`. This goes in the direction of unique definition, as a general concept such as “development” always has its own notation `-027.1`, which can be attached to any other class:

`338.48-027.1(410-15-22)` tourism development in rural western Scotland

11: A one-table scheme

Both in the CRG draft scheme and in BC2, attributes are listed in the schedules before entities, according to the principle of increasing concreteness. Entities follows, in order of integrative levels, starting from simple physical particles, and going on to molecules, cells, organisms, etc. This sequence reminds very much that of the Roget's Thesaurus in the outline of the 1911 edition, recently revitalized to be used as an access point in the Wikipedia [2007]. Indeed, in the prototype of the Thesaurus the first listed terms are “words expressing abstract relations” (`I`), followed by “words relating to space” (`II`), “words relating to matter” (`III`), etc. In a modern form, this gives the following classification system, looking not far from the idea of a single general scheme arranged by integrative levels:

```
1       abstract relations
11          existence
12          relation
13          quantity
```

14 order
15 number
16 time
17 change
18 causation
2 space
21 space in general
22 dimensions
23 form
24 motion
3 matter
31 matter in general
32 inorganic matter
33 organic matter
332 sensation
3322 special sensation
33221 touch
4 intellectual faculties
42 communication of ideas
423 means of communicating ideas
4232 conventional means
language
letter
word
...
5 voluntary powers
6 sentient and moral powers

It can be seen that, in this scheme, notions like time, space, and form are not managed separately as common auxiliaries, but are in the same main sequence as the more "concrete" entities, according to a unique logic. (The expected use of terms in the original thesaurus was obviously different from that of modern classification schemes, but this does not affect our present point.) Such a solution agrees mostly with the principle of unique definition of a concept: each concept has a single place of definition in the schedules, and is assigned a single notation. All the needed combinations can then be made possible, but the rules of combination are clearly separated from the table of the concepts.

To summarize, the following devices can exist in a classification scheme:

- *common auxiliaries*, or *common facets.* They are facets that can be applied to any basic class throughout the whole scheme, irrespective of its disciplinary context. In an expressive notation, they are introduced by a facet indicator (punctuation mark in UDC; digit in BC2; letter in FATKS; one or more digits in ILC) followed by the *foci.* These can be linked to unique definition of main classes (FATKS), or even be borrowed from them (ILC);
- *facets.* Unlike common auxiliaries, they are applied only within a special class or subclass,

although the mechanism is similar. Their foci can be linked to, or borrowed from, classes in the main table. In some cases, however, foci have a meaning only within that facet: gender is only a facet of organisms, and its foci (male, female) cannot be taken from elsewhere in the scheme; in other words, their place of unique definition is that facet itself. In ILC these are called *context-defined foci*;

- *attributes*, like properties, processes and operations. In faceted schemes they are treated as facets, either common (for general attributes appliable to any concept), or specific of a class (for attributes defined only in relation to a specific concept);
- *phases*. They can be any class from the main table, and are connected by phase relation operators;
- *subject device*. It can be used in some classes that are potentially related with any other phenomenon: "education" can have anything as its subject; "libraries" can be specialized in any topic. In ILC they are called *general extra-defined foci*, and work as facets. In these cases, their notation is taken entirely from that of the phenomenon in the main tables, without alterations. This is done also in enumerative classifications, although not always defined explicitly;
- *dependence* and other associative relatioships. They are not expressed in notation, but can be recorded in the schedules, to mean that a concept is connected to another one;
- *discipline-phenomenon* relation. In disciplinary classifications, phenomena are defined only as the object of one or more disciplines; there can be classes for phenomena treated in an interdiscipinary way: in this case, there should be a link between each phenomenon and its definition as the object of a class, e.g. between "plants" and "botany". In a phenomenon classification, the perspective is reverted: classes are made up of phenomena, and disciplines are seen as a particular class of phenomena at the integrative level of human culture; the studied phenomena are thus simply a facet of the discipline, and their values are borrowed from the appropriate level of phenomena:

```
M           organisms
Mp                  plants

Yn          sciences
Yn8 [*]             sciences studying objects...
Yn8mp                       sciences studying plants, i.e. botany
```

In the interest of both internal consistence and information retrieval effectiveness, all these devices should be explicitly connected to the place of unique definition of the related concepts.

Acknowledgments: Many thanks to Ia McIlwaine for her willingness to provide advice to improve a previous version; to Aida Slavić for her friendly and qualified help in clarifying details of classification schemes; and to Caterina Barazia for nicely agreeing to include this conference in our wedding travel...

References

Austin, D. 1969. Prospects for a new general classification. *Journal of librarianship*, 1, n. 3, p. 149-169.

Beghtol, C. 1998. Knowledge domains: multidisciplinarity and bibliographic classification systems. *Knowledge organization*, 25, n. 1-2, p. 1-12.

Blair, D.C. 1980. Searching biases in large interactive document retrieval systems. *Journal of the American Society for Information Science*, 17, n. 3, p. 271-277.

Broughton, V. & Slavić, A. 2007. Building a faceted classification for the humanities: principles and procedures. *Journal of documentation*, in press.

Classification Research Group. 1955. The need for a faceted classification as the basis of all methods of information retrieval. *Library association record*, 57, p. 262-268. Republished in: *Proceedings of the international study conference on Classification for information retrieval, Dorking (England), May 13-17 1957*. Aslib: London, 1957, p. 137-147. Republished in: Chan, L.M. (Ed.). *Theory of subject analysis: a sourcebook*. Libraries unlimited: Littleton, 1985.

Farradane, J.E.L. 1952. A scientific theory of classification and indexing: further considerations. *Journal of documentation*, 6, n. 2, p. 83-99.

Foskett, D.J. 1970. *Classification for a general index language: a review of recent research by the Classification Research Group*. Library Association: London.

Gnoli, C. 2005. BC2 classes for phenomena: an application of the theory of integrative levels. *Bliss classification bulletin*, 47, p. 17-21. Also in *DLIST*, <http://dlist.sir.arizona.edu/920/>.

Gnoli, C. 2006. The meaning of facets in nondisciplinary classifications. In: Budin, G., Swertz, C. & Mitgutsch, K. (Eds.). *Knowledge organization for a global learning society: proceedings 9th international ISKO conference, 4-7 July 2006, Vienna*. Ergon: Würzburg, p. 11-18.

Gnoli, C., Bosch, M. & Mazzocchi, F. 2007. A new relationship for multidisciplinary knowledge organization systems: dependence. In: Blanca Rodríguez Bravo, M.a Luisa Alvite Díez (Eds.). *La interdisciplinariedad y la transdisciplinariedad en la organización del conoscimiento científico = Interdisciplinarity and transdisciplinarity in the organization of scientific knowledge : actas del VIII Congreso ISKO-España, León, 18, 19 y 20 de Abril de 2007*. p 399-409 León: Universidad de León, p. 399-409. Also at <http://isko2007.unileon.es>.

Gnoli, C. & Hong M. 2006. Freely faceted classification for Web-based information retrieval. *New review of hypermedia and multimedia*, 12, n. 1, p 63-81.

ISKO Italia. 2007. *The León manifesto*. <http://www.iskoi.org/ilc/leon.htm>.

McIlwaine, I.C. 1997. The Universal Decimal Classification: some factors concerning its origins, development, and influence. *Journal of the American Society for Information Science*, 48, n. 4, p. 331-339.

Parsons, E.C.M., Warburton, C.A., Woods-Ballard, A., Hughes, A. & Johnston P. 2003. The value of conserving whales: the impact of cetacean-related tourism on the economy of rural West Scotland. *Aquatic conservation: marine and freshwater ecosystems*, 13, p. 397-415.

Perreault, J.M. 1965. Categories and relators: a new schema. *Revue internationale de documentation*, 32, p. 136-144. Republ. in *Towards a theory for UDC*. Bingley: London 1969, p. 119-140. Republ. in *Knowledge organization*, 21: 1994, n. 4, p. 189-198.

Ranganathan, S.R. 1961. Library classification on the march. In Foskett, D.J. & Palmer, B.I. (Eds.). *The Sayers memorial volume*. London: Library Association, p. 72-95.

Ranganathan, S.R. 1967. *Prolegomena to library classification*. 3rd ed. SRELS: Bangalore. Also in *DLIST*, <http://dlist.sir.arizona.edu/1151/>.

Slavić, A. 2006. Interface to classification: some objectives and options. *Extensions & corrections to the UDC*, 28: 2006, p. 24-45. Also in *DLIST*, <http://dlist.sir.arizona.edu/1621/>.

Slavić, A. & Cordeiro, M.I. 2004. Core requirements for automation of analytico-synthetic classifications. In McIlwaine, I.C. (Ed.). *Knowledge organization and the global information society: proceedings of the Eight international ISKO conference, 13-16 July 2004, London*. Würzburg: Ergon, p. 187-192.

Sukiasyan, E. 1998. Classification systems in their historical development: problems of typology and terminology. In: Mustafa ed Hadi, W., Maniez, J. and Pollitt, S.A. (Eds.). *Structures and relations in knowledge organization: proceedings of the Fifth international ISKO conference, 25-29 August 1998, Lille*. Ergon: Würzburg.

Szostak, R. 2007. Classification, interdiscipinarity, and the study of science. *Journal of documentation*, 63, in prep.

Tomlinson, H. 1969. Report on work for new general classification scheme. In: Classification Research Group. *Classification and information control*. Library Association: London, p. 29-41.

Vickery, B.C. 1960. *Faceted classification: a guide to the construction and use of special schemes*. Aslib: London.

Wikipedia, upd. 2007. *Wikipedia: Outline of Roget's Thesaurus*. <http://en.wikipedia.org/wiki/Wikipedia:Outline_of_Roget%27s_Thesaurus>.