Jacques Mauiez
University of Dijon (France)

A Decade of Research in Classification

Maniez, Jacques : **A decade of research in classification.**
Int.Classif. 18(1991)No.2, p. 73-77, 8 refs.
The seven reports written by Eric de Grolier for the FID/CA Committee 'General theory of classification' between 1953 and 1960 are a precious testimony of the author's reflexion and methodology, and also of the state of classification issues in the fifties. The main content of these reports is a general and evolutive project devised as a basis for a new universal standard classification. An original type of alphanumeric and pronounceable symbolization is advocated which would allow a flexible division of main classes between the domains of knowledge. Besides he reviews the emerging documentary languages (thesauri, key-words) and the new automatic retrieval devices. De Grolier's studies are based on an impressive erudition and a prudently experimental approach. (Author)

0. Delimitation of the study

In 1953 Eric de Grolier was entrusted by F. Donker Duyvis with writing a preparatory report for the FID Committee 'General theory of classification', and he continued to fulfill this task through the early sixties. He wrote a total of seven reports, that we have recently become acquainted with, except for the sixth one, which could not be found[1].

The scope of the committee, as defined in the first report, was broad and ambitious: "*Documentary taxilogy* will not grow into a science until it brings into confrontation, without any prejudice, data and theories, established facts and hypotheses" (I, p.105). Consequently many different questions are dealt with in the reports. We shall focus on the main issues: symbolization, contents and structurization.

1. Symbolization in encyclopaedic classifications

Among the hundred pages that make up these reports, nearly fifty are devoted to the problems of symbolization. In fact it was then commonly admitted that the choice of a decimal notation by the founders of the UDC had led to a dead end, and that other solutions were to be found. Moreover new tools for automatic information retrieval were tested, some mechanical, others electronical, and at that time many people assumed that concepts and class-numbers ought to be coded according to the requirements of the machines.

In that context it was worthwhile going deeper into the problems of symbolization, and many prototype coded languages were developed in the fifties (John Melton's 'Semantic code', Robert Pagès's 'coded analysis'...).

1.1 The principles

Eric de Grolier himself played a major role in this area of research. He explains his problematics in the English summary of his first report (I, p.116):

1) "There is a tendency towards alphadecimal symbolization... but it is necessary to choose... the letters of the Latin alphabet which are in the same order when transliterated into cyrillic characters".

2) The chief advantage of letters over Arabic numerals is phonetical, in so much as "they have an international standard value ... and are much briefer, when pronounced".

But precise rules should be observed in order to limit the letters and the combination of letters to sounds that can be pronounced in the same manner by all people of the world, whatever their language may be.

In short two preferential ways are assumed: the use of alphabetical symbols and the choice of pronounceable sequences. But each way is severely restricted when you go from a national and one-language to a universal and all-language context.

1.2 The alphanumeric solution

De Grolier's option of an alphanumeric solution and of the Roman alphabet was firmly settled from the outset and remained unchanged; since Chinese linguists had adopted Roman characters to transcript the sounds of their language (pinyin), the Roman alphabet was de facto an international standard. However some letters, the order of which is different in the closely related cyrillic alphabet, should be eliminated, so that a sequence of 17 letters was left available: (a,b,d,e, i,j,k,i,m,n,o,p,r,s,t,u,y). That set was the basis of the system, and the figures had to play a complementary role, for instance in briefly coding the names of countries or the format of documents.

1.3 A pronounceable and ordered symbolization

On the other hand, the phonetical quality of the notation was a feature that de Grolier felt strongly devoted to, but raised difficult issues. He took advantage of his enormous knowledge in linguistics and phonology to perfect a truly international symbolization system, but the critiques of his colleagues and his own scientific scruples often led him to repentance. Here are, roughly explained, the successive schemes that he put forward within that period.

In the first report he points out that a sequence of letters can be easily memorized when one can utter it without effort (lap, irma, put ...), while it is harder to remember if it is impossible to pronounce (wtzs, lfvg ...). He also asserts that "oral communication is developing more and more with the expansion of the magnetic recording of sounds ..." (I, p.106). But the requirements of an international standard greatly limit the set of authorized sequences in so much as many combinations of phonemes can work only in some languages and many sequences of letters are pronounced differently in the different countries.

To fulfill these requirements he follows the phonetic rules prescribed by the Russan linguist Trubeckoj and advises to restrict the use of letters to those which have a contrastive character to others in every language. In such conditions the standardized notation could be made of only 12 letters, to which a conventional phonetic value would be assigned (a,e,i,k,l,m,n,o,p,s,t,u). In a future universal classification, each hierachical level of division could be symbolized either with a letter, or with a syllable of two or three letters, which would afford a sufficient set of posssibilities. Some syllables could keep a constant semantical value, like the morphemes in natural languages, and thus be used in the formation of complex symbolic sequences. As for the relational symbols, a whole of five symbols would be enough, which could be expressed by five distinct unused syllables.

In the next report, de Grolier reviews the criticism his first scheme had received, and he admits that a severe restriction of the allowed characters tends to lengthen the sequences of symbols. Which then should be sacrificed? The capacity of the standardized notation or its practical universality? After a detailed survey of several alternatives de Grolier strays a little from the rough principles of Trubeckoj's model and proposes a larger list of allowed combinations of letters (147 two-letters syllables, 1679 three-letter syllables).

The fourth report, entirely devoted to symbolization, widens the field of phonetic solutions by proposing a new scheme, the 'phono-centimal' system. That scheme takes into account only the figures and gives a syllabic equivalent for each of the 100 combination of two figures (for instance 'pa' for 10, 'pe' for 11 ...). The originality of this project is that it is based on a learned and systematic study of the phonemic structure of language. Inspired by recent works on the distinctive features of phonemes, de Grolier proposes to replace the traditional order of letters with a more natural order, which could make it possible to assign related sounds to related fields in a classification.

The next report, published shortly after the International Dorking Conference (1957), and called significantly 'After Dorking', shows the signs of some changes in de Grolier's conception of symbolization. Indeed he keeps on striving for an alphabetical and phonetical notation built upon logical and scientific bases, which he expresses in the manner of Ranganathan's postulates (V, p.15):

"The basis for symbolization is the set of small letters of the Latin alphabet ... with a fixed phonetical value and in an order determined by the distinctive oppositions of the corresponding phonemes ...

- The combinations of letters ... are adjusted so that the coded-words can be pronounced clearly and unequivocally.

- The distribution of symbols ... endeavours to allot the shortest notation to the most-used words

- The structure of symbolization does not necessarily correspond to the hierarchical organization of symbolized notions ...".

But on the other hand he clearly questions his previously firm belief in the 'raison d'être' of an ordered symbolization. Stating that this domain is the one in which the least agreement was reached during the conference, he calls for a critical and autocritical approach: "Perhaps till now we have not been attentive enough (and I am the first to blame) to the fact that coding for machine language leaves out a formerly primordial necessity in traditional classifications: that of assigning a conventional order to the symbols. The machine language needs only strictly defined coded words ... That makes a part of our previous research purposeless ..." (V, p.10).

Though the last report (1960) brings few new elements to the 1958 report, it tackles again and refines some aspects of the previous scheme. De Grolier gives up on establishing a relationship between groups of phonemes and fields of knowledge, and restricts the number of authorized syllables.

But in other respects he regrets that the new order of letters he had devised no longer reflects for the user the parallelism between the contents and the codes, and he proposes a numerical extra-code reflecting this logical order (VII, p.36). And, looking back on his previous projects, he betrays a little disillusionment "In our former reports we have tried to put in agreement a classification of *logically* ordered phonemes and the classification of symbolized concepts. We believe it is honest to confess that none of these trials was very satisfying" (VII, p.32). A similar feeling of disappointment emerges from the *Study on general categories* ... , which was written in 1959, but published only in 1962 by UNESCO. In that book of 262 pages, the problems of symbolization

are relegated to a last chapter of only one page, and the author explains his reasons "I intended, in a first stage, to examine the problems of symbolization ... , about which I had already published some preliminary studies,-which are today outdated-.... Those questions should be taken up again on an entirely new basis ..." (1).

As a conclusion, this part of de Grolier's reports may sound disappointing by the drastic - and far too strict - criticism the author applies to his own work. We feel, on the contrary, that it gives us evidence of the author's intellectual probity.

The extent of de Grolier's scientific background is impressive, especially in the fields of epistemology, philosophy and language. Before writing on a topic he makes it a point of honour to have read whatever relates to it, and quite technical features are enlightened by references to basic sources: Jakobson, Piaget, Trubekkoj ... Then, with passion, he aims at achieving a synthesis of all the best of the existing theories. But this ambitious goal is difficult to achieve, and sometimes frustrating, as witnessed by de Grolier's successive propositions, retouches, adjustments and final repentance, regarding the creation of a new symbolization. However he always proves ready to start again "on a new basis" (1).

2. The contents of a modern classification

This core question fills nearly one third of the six reports, and there de Grolier fully shows the capacities of his encyclopedic mind.

Starting from a critical view of Bliss's theory on the 'consensus' of learned people as a basis for classification designers, he argues that the system of sciences, in spite of continuous revisions and contradictions, tends toward a state of equilibrium and that in the 1950's scientific convergences prevail over divergencies.

First he proposes to distinguish the relatively static domains (philosophy, religion, social sciences) from expanding and dynamic domains (mathematics, physics, chemistry and biology) and to divide the main classes accordingly , while keeping in mind that empty slots should be reserved for any future expansion.

Then he studies the conditions required to build a truly universal classification, and as a means of measurement he compares the relative importance given to the different countries, religions and languages in five classifications (Dewey, UDC, LC, Bliss, CC).

It appears that a new classification should keep away - as far as possible - from the temptation of ethnocentrism, and with that aim in mind, should be designed by a panel of scientists from all continents.

These principles led to a tentative scheme called ALSYN (ALphabetical and SYNthetic), which was but slightly revised in the following years. Here is a table of the second version (54), which is made up of 12 main classes (the figures show the average percentage assigned to each class in the notation).

Logic, dialectic, mathematics	10%
Physics	18%
Chemistry	7%
Cosmology	7%
Biology	6%
Botany and zoology	5%
Physical anthropology	10%
Sociology, linguistics, history and geography	7%
Technology and economy	15%
Politics, law, morals, education, philosophy	8%
Arts and games	5%
Literature	7%

Unlike Dewey Classification, in which the main classes agree with the main disciplines of the nineteenth century and the order of classes is not really significant, the domains in this scheme are rather areas of interest, and their sequence reflects a logical view : from abstract and general surveys to concrete and complex things, from inanimate nature to live world and man. Moreover, as we have seen before, the alphabetical notation allows for the adjunction of new basic classes.

De Grolier's scheme does not intend, in this first stage of the project, to go beyond the array of the main classes, because this level is the basis for all the further evolution of the system, and it would be irrelevant to go further before reaching a general agreement on it. Nevertheless, interesting views on the organization of some domains are to be found in two of the reports.

The second report deals with the possibility of a unifying principle between four related and interrelated fields: geography, ethnology, history, linguistics. Relying on recent theories, the author pleads in favour of a parallel arrangement, based upon a geographical partition:

The old world (Africa, Eurasia, Europe, Asia)
The new world (North and South America).

The seventh report briefly reviews the classification systems of some fields of science and technology (logic, chemistry, engineering, lexicology). But here the methods seem to be more specific than convergent.

3. Theory of classification

First, let us stress that at this stage of his work, de Grolier, like many others, uses the word 'classification' both in the general sense of 'documentary language' and in its classical and narrower sense. (Ten years later, in his masterly lecture on *'The system of sciences and the evolution of knowledge'*, at the 1971 Ottawa Conference (2), he was to carefully make the distinction).

The starting point of the author is clear-cut: "The theory of classification has progressed significantly over

the past twenty years ... The main object of our research should now be to integrate these new contributions ... into the Otletian legacy" (I, p.114). In other words, de Grolier's assumptions could be summarized as follows:

- the main intuitions of the UDC designers were practical (multi-coordinated classification, adequate syntax of few symbols)

- unfortunately these principles have not found a relevant application within the restrictive framework of Dewey's decimal system

- consequently a new standardized international classification is needed. And the most urgent task lies in proposing a new basis for the sharing of knowledge and correlatively a new system of symbolization.

As we have seen, this task occupied the greater part of de Grolier's studies during that period, and the six reports assign only about twenty pages to theoretical issues in classification: two reports deal with some basic issues of the inner structure of the project, while the last one reviews new tendencies in 'documentary classifications', namely in indexing languages. Let us comment briefly on each point.

3.1 Parallel or autonomous division of the main classes?

It is well-known that this crucial point determines the structure of a classification : the principle of autonomy leads logically to a fully-hierarchical structure, as in the LC, while the principle of parallelism, when carried to its extreme limits, leads to a faceted structure, as in the CC. Of course, many intermediary solutions may be tried, among which we find a partial parallelism in the separation of related domains (ethnology, geography ...), or the mere transfer of a division scheme from one part to another with the formula "Divide as for ...".

The first report of 1953 comments on those intermediary solutions. De Grolier firmly states the theoretical advantages of parallelism: "Indeed an encyclopaedical classification of knowledge should not be ... a mere mosaic of juxtaposed specialized classifications. On the contrary it should attempt to establish a coherent system" (I, p.113). Yet, dealing with the process of transferring a type of division from one class into another (as is usual in the UDC or the CC) he proves to be very cautious in applying this device to interrelated classes such as linguistics, geography, ethnology, religion. And one year later, he proposes, in a more prudent approach, to adopt the same point of view (geography) and the same order in dividing those fields, but while adapting this general scheme to the specificity of every domain.

As for the extreme and systematical solution of facets, de Grolier was sometimes tempted by it, but in the end he remained distrustful. Indeed in the third report he acknowledges the fruitfulness of such research (III, p.15) and he proposes his own list of viewpoints which could be applied to all domains: methodology, forms, functions, types ... But unlike Ranganathan's facets, which aim at subdividing every field in an attempt to achieve the canonical expression of all potential subjects, these categories are only general viewpoints adjusted to the Aristotelian principle of division.

In the same way, it seems that he agrees with a faceted structure in 1958 when he outlines the general postulates of his scheme (V, p.15):

"- Inside every domain ... the analysis of concepts to be codified should be made by 'categories' ....

- Complex notions should be analysed into codified elements and then recombined by synthesis."

But in the same report he thoroughly criticizes Ranganathan's PMEST facets (V, p.1-5), demonstrating with numerous examples that these categories - or even Vickery's more logical categories - cannot be easily adjusted to such domains as linguistics, art, philosophy or religion. And in the last report (VII, p.4-6) he insists that, while faceted classifications are expanding in some narrow fields, they do not seem likely to create a new encyclopaedic classification in place of the UDC.

From those statements we may reasonably assume that during that period de Grolier finally remained devoted to a hierarchical structurization of the main classes in a universal classification, except in some particular fields, but that he advocated using the device of parallel divisions in order to ensure the coherence of the tree-structure.

3.2 New documentary languages

De Grolier devotes a large part of his last report (VII, p.17-28) to the topic "Linguistics and classification" (the term 'classification' is used here in the broad sense of 'documentary language').

Commenting on recent work in automatic translation, and chiefly on Bar-Hillel's theories, he insists that such research as language processing by computers, semantical analysis, thesaurus approach and new methods of information retrieval are likely to bring big changes in libraries and information centers.

Obviously, he observes that tendency, (which at that time was about to generate a new and prevailing type of indexing languages) with interest and sympathy, but his statements remain prudent and he does not raise the basic question whether conventional classifications are still relevant to building a new universal documentation scheme.

4. Some concluding remarks

Four decades later, these reports remain a precious source of knowledge and reflection. They are clear evidence of de Grolier's background and methodology, and more generally of the history of information science in the fifties. Moreover, a large part of their content is still relevant.

4.1 De Grolier's scientific methodology

We have already insisted upon the encyclopaedic background and the intellectual probity of this researcher. Another striking feature lies in his constant devotion to experiment as the last evidence of scientific truth. His readings are widely extended in all fields and in several languages, and he is fond of getting acquainted with new theories, while always keeping a critical and cautious view of them, and he feels sceptical towards intellectual models that are too systematically organized. In that respect his relationships with Ranganathan's theory seems to be exemplary. While Ranganathan had a mathematical background and consequently tended to believe in a logical basis to human knowledge as a whole, de Grolier, who chiefly specialized in the social sciences and linguistics, was influenced by the diversity and the specificity of the different fields. As we have seen, in the crucial debate on the ideal model of a universal classification between the advocates of a symmetrical structurization of the main domains and those who favour autonomous divisions, he first tried to conciliate both approaches, but ultimately preferred the hierarchical structure because it allows each field to keep the degree of autonomy it needs.

4.2 Forty years later ...

Of course, time, history, society and information have evolved along lines that were then unforeseeable. The UDC has somehow proceeded on its merry way and no alternative scheme has contested recognition; MacLuhan's predictions have been proved wrong, and libraries - whether automated or not - are prosperous ; indexing languages are now prevalent, but they have not superseded classification schemes. But in spite of so many changes, a good deal of de Grolier's observations and projects retain a theoretical and practical value today. His scheme of an utterable notation - however strict in its form - affords an original basis for user-friendly symbolization in any new classification, either special or universal.

We think that ALSYN classification, forty years after its conception, could still be a satisfying alternative to the main classes of the UDC in so much as it does away with its main drawbacks, namely its anachronism, the insufficient room devoted to dynamic fields, the fact that it is a closed system, the separation between science and technology. The new disciplines (e.g. computer science, genetics, generative linguistics) could easily be incorporated. The ordering of its classes is logical and aims at achieving the collocation of related fields. (Besides, the UNISIST's Broad System of Ordering - 1978 - shows some degree of similitude with ALSYN).

Finally, his views on the structuring of classification systems are still topical because they constitute the basic questions and offer balanced solutions.

Notes

1 In the following quotations, every report is referred to by its Roman numbers. The translation from the French is ours.

References

E.de Grolier's reports

I L'étude des problèmes de classification documentaire sur le plan international. (with English summary) Rev. Doc. 20 (1953) p.105-117

II Problèmes de classification documentaire sur le plan international. FID document F54-81 (1954) 21 p. + 7p. appendix. Mimeographed

III Nouvelles recherches sur la symbolisation des classifications documentaires. Rev. Doc. 23 (1956) p. 13-21

IV Symbolisation normalisée d'une classification encyclopédique internationale. Rev. Doc. 23 (1956) p. 128-139

V Après Dorking. FID document 025.4(04) (1957) 16 p. Mimeographed

VII Tendances actuelles en matière de classifications et codifications documentaires (with English abstract). FID document (1960) Mimeographed

Other References

(1) Grolier, Eric de: Etude sur les catégories générales applicables aux classifications et codifications documentaires. Paris: Unesco 1962. 165p.

(2) Grolier, Eric de: Le système des sciences et l'évolution du savoir. In: Conceptual basis of the classification of knowledge. Proceedings of the Ottawa Conference 1971. München: Verlag Dokumentation 1974. p.20-118

Address: Dr.Jacques Maniez,
61, rue de la Corvée, F-21000 Dijon, France