ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

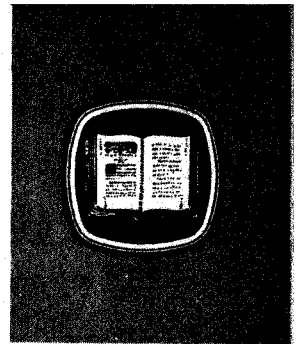University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

RE-CLASSIFICATION: SOME WARNINGS AND A PROPOSAL

by Jean M. Perreault

Summary of Contents

ABBREVIATIONS

ALA	American Library Association
APUPA	Alien, Penumbral, Umbral, Penumbral, Alien
BC	Bliss's Bibliographic Classification
CC	Colon Classification
DC	Dewey Decimal Classification
HEA	Higher Education Act of 1965
LC	Library of Congress Classification
UDC	Universal Decimal Classification

4

0.1 Prelude

Classification can be the arrangement of documents on the shelf; it can also be the arrangement of the surrogates of those documents in the catalog. Subject-headings can hardly be the first, though they certainly are the second. But forgetfulness of the purpose of all these modes of arrangement (namely, their function as strategies for the searching-out of documentary information either known or unknown to be available) has led to non-application of over-arching guiding principles in the arrangement of both catalogs and shelves. When these principles are once more brought to consciousness, we can establish criteria for the judging of all varieties of search strategies; when they are not, criteria of wrong sorts must instead come into play, leading to disastrous results both in terms of quality of service and of expenditure of funds.

That the disastrousness of these results is not even noticed is a perfect demonstration of the lack of adequate criteria and of the principles in which they originate. The current rush to re-classify to the Library of Congress classification (LC) is one example of such a result stemming from criteria of the wrong sort--criteria arising in the absence of guiding principles. Such warnings are made here in terms of the presence or absence of such principles, criteria, and operations; but a proposal is also made: to re-classify to the Universal Decimal Classification (UDC). This proposal, to be as effective practically as it is theoretically, must imply the establishment of logistical means to compete for the market with LC, etc.; but there are serious hurdles to be overcome in such an establishment.

Re-classification is a problem accentuated by the present movement to automate general libraries; such automation should call forth re-examination among librarians as well as documentalists (whatever the distinction, if any, except in terms of willingness to re-examine) of what our fundamental purposes are and of how we can best meet them. It is often, unfortunately, only the wedge with which to insert un-reflective re-classification into the context of total change. Automation can lead--and perhaps must lead--to new ways of meeting old purposes as well as new; what follows should be taken as referring to the automation situation even more than to libraries still conventional at least in the means of service.

1.1 Occasions, of Various Sorts

American librarians can hardly be unaware of a galloping tendency, whenever any particular institution begins to fear that its mode of internal arrangement is inadequate, for the response to be something like: "All right, then, we had better re-classify to LC." This is so, whether because of unprecedented growth, because of growing demands for utilization--with a concomitant pressure for the release of energy and money from the non-public-contact sectors of its work--or because of dissatisfaction with the mode of internal arrangement in use.[1] Why this response is so predictable is one part of what will be discussed below; why it is unfortunate in the extreme, together with what could be done as a superior alternative, is another part. It would be interesting to speculate as to whether any libraries now using LC, and in a position of saturation analogous to that of libraries now planning to change from the Dewey Decimal Classification (DC) to LC, have any clear idea of how they can escape from the analogous difficulties.

Three recent personal contacts have been additional occasions for the present essay: with Phyllis Richmond, who, in her address "General Advantages and Disadvantages of Using the LC System,"[2] attempts to set down a foundation of more than mere expediency for the tendency to the automatic pro-LC response;[3] with Jerrold Orne, who appeals primarily to fiscal considerations (and to a strong scepticism about the value of classification as anything more than a "finding device") for the decision to go from DC to LC at the University of North Carolina; and with Harry Dewey, who points out the paradoxicality of the simultaneity of this tendency and of the reawakening of interest in classification theory in America.

But unquestionably the largest reason for the automatic tendency is the enactment of the statutory means for the Library of Congress to eliminate the large numbers of foreign documents which, in the recent past, had constituted so formidable a proportion of the work of the larger and/or more specialized libraries.[4] If there had been a tendency for libraries under pressure of growth to change to LC even before the enactment of the Higher Education Act of 1965[5] the pressure to give in would become all the more tempting when one hoped that one might thereby never need catalog again for oneself, thus releasing energy for "public service" and money for collection building.

1.2 Some Arguments in Favor of LC

There are, of course, arguments in favor of LC; otherwise no one would be tempted by it. Richmond's central contention is that it is flexible because it is non-systematic, and, in the legitimate sense that the proof of the pudding is in the eating, that it simply works. Orne's is that in the absence of any overwhelming conviction of the superiority of any other classificatory system over it,[*] one naturally tends to choose the system that saves money. Cronin's is that international exchange of bibliographical data is a good thing, and that if others besides the Library of Congress can profit by such cooperation, so much the better.

Richmond's central contention is based on an argument which seems to be wholly self-destructive, because it postulates a total lack of detectable or (even more important) predictable system in LC. I do not agree with so harsh a judgment against LC; but in her mind there is nothing harsh about it, since the (non-)system works anyway. Perhaps these contradictory points can be resolved in a clarification of what is meant by "it works"; it is not clear what Richmond means by this when applied to a classification, but what the present writer means by it, will be explored further in §2.1.

Richmond's most compendious statement of her thesis reads:

> One does not have to look for a subject where it <u>ought</u>
> to be. One only has to find it where it is ....
> One may achieve a great degree of consistency in
> classification if one does not have to fit new ma-
> terial into a logical pattern, but only into either
> a precedence pattern or into loopholes in an ordinal
> system.[6]

---

[*]That even if this argument be correct it leads to fallacious further conclusions will be shown in §2.2.

6

Orne evinces a thorough scepticism as to the demonstrable superiority of one classification over another, LC included; the ineluctable conclusion must then follow, other things being equal, that the classification which costs least to use is to be used. I do not agree that there is really nothing to choose among classifications, nor that (except Orne's opinion be demonstrated to be true) such considerations as he puts forward are professionally acceptable.

Cronin's point is one that cannot be so easily attacked, because its presupposition is one which none could reject, namely that work done once for all is preferable to the same job done over and over without any progressive improvement. What is really essential in the international exchange of catalog-data that Cronin describes are the nominal and formal elements, not the conceptual;[7] it was to these first two elements that the Paris Conference directed its attention--and it is from this conference that have emanated the influences which have resulted in the Anglo-American agreement described by Cronin. But there is no impression in Cronin's article that LC classification is expected to conquer the British library world.

1.3 Experience with LC at Florida Atlantic University

The papers "On Bibliography and Automation"[7] were directed to the problems of the interaction between the exigencies of automation and the principles of bibliography, particularly of conceptual bibliography; the general conclusion was that only if the latter are fully understood and adequately implemented, and only if the machine exigencies are so modified and modeled as to conform to these more fundamental goals and strategies, can what is achieved be prevented from becoming a white elephant of massive proportions and shocking cost. The mentioned papers were occasioned by both an external and an internal pressure. The external one was brought to bear mostly by Ellsworth's paper cited above[4] which, even before the enactment of HEA, seemed to suggest that the member libraries of the Association of Research Libraries were ready to accept almost any way out of their overwhelming expensive burden of cataloging over and over, among them as a group, what was not centrally available.

The internal pressure was that occasioned by the situation at Florida Atlantic University, where, the catalog having been successfully automated, need arose to consult the electronically stored information. An order had been given for an exhaustive search preparatory to the establishment of an area-study program on the Caribbean. Two possibilities were available: consultation of the classification codes, or consultation of the subject-headings. Unfortunately, neither proved to be workable as strategies for consulting the electronically stored conceptual-bibliographical information, and therefore there had to be made a manual search of the whole of the catalog in order to guarantee exhaustiveness. The conclusion was inescapable, given at length in "On Bibliography and Automation,"[7] that there should have been a far more thorough preparatory examination of the potentialities of such catalogs, including an attempt to determine whether there was a substantive difference between the various available general classifications, one which would enable more to be done with the electronically stored catalog records than to substitute computer for typist, compositor, or filing clerk. For if these functions are all that are going to be expected, extreme problems will assuredly arise with the cost of cumulation of the resultant printed catalogs.[8]

The classification codes for the Caribbean area (and, of course, any of its included parts),[9] extend from F1601 to F2175. It would be a very simple matter to have the computer print out whatever nominal and formal elements were desired for all documents which had had any such codes assigned them. Unfortunately, though, such a search would still be a very long way from exhaustive, since the mentioned span of codes relates only to documents in which the Caribbean area is treated geographically or historically; documents on education, art, government, biography (or any other primary headings under which a "relative" such as the Caribbean area might be "distributed"), would be found only by searching though the appropriate spans of codes, and then by accepting only those codes to which had been affixed the alphabetical or tabular code representing the Caribbean area or its parts. Thus:

LC1071 (examinations in professional schools) is divided alphabetically by country, so that for Cuba it becomes LC1071.C8;

LE15-17 is individual educational institutions in the West Indies, semi-enumerated alphabetically "by city, province, island, etc." in the main tables;

N910 (art galleries in the Western hemisphere outside the U.S.) is divided alphabetically by city, so that for Havana it becomes N910.H3;

NE501-794 (history of engraving) is divided systematically by area and then alphabetically by country in Table III-A of the N-schedules, giving NE568-586 for the West Indies;

JA84 (history of political science by country) is divided alphabetically, so that for Cuba it becomes JA84.C8;

JL590-1169 includes all the West Indies as related to constitutional history and administration, but also many irrelevant places, so that a precise search would have to accept only JL590-669, JL740-779, JL790-799, JL820-840, and JL1000-1169;

JS1851-2059 is local government in the Caribbean area, enumerated in the main tables (paradoxically, it does not include the irrelevant Guianas, as found in JL590-1169);

CT330-548 includes all the West Indies as related to biography, but includes the same irrelevant places as we saw in JL590-1169, so that a precise search would have to accept only CT330-398, CT430-448, CT470-498, and CT510-548.

It may not be expected, for one who is familiar only or primarily with LC, that there be any such predictable similarity among these various correlated codes which would make them recognizable by a computer without extensive and expensive re-programming for each subject for which there could be a correlation with the desired idea. But such a lack would not be acceptable to those who see the possibility of a positive solution of the problem.

The other possible line of approach was through the subject-headings. Here we might expect that we could start from the term "Caribbean area" or "West Indies"

and proceed thence, via the see-also references, to the locating of all the included terms. But first it is necessary to remember that the see-also references in LC subject-headings (and in Sears as well, of course, based as it is upon LC) are not thoroughly organized to provide chains from broadest to narrowest, especially when the terms are proper names. Nor are LC "see-also" references unambiguous; they indiscriminately refer to subordinates, coordinates, correlates, and even occasionally to superordinates. In addition, there is the phenomenon described in an unpublished paper by Ritvars Bregzis, "Automation and Bibliographic Control," of "loops" in the "see-also" reference structure, which would, unless prevented by program-error-detection, cause trouble in computer searching. The first of these problems is the really serious one in the case at hand, as can be seen from such a heading as "Mailu," a tribe in New Guinea which is not connected by see-also references to any including or included headings, either ethnic or geographic. We cannot then expect to make an exhaustive search for New Guinea (and the same is true of any other geographic area) unless a preliminary search be made for all the potentially productive headings not connected by "see-also" references. Thus the computer search of LC subject-headings was abandoned, since it would be successful only if a manual search were made first.

1.4 The Demands of Professionalism

The reasoning that leads to the decision to re-classification to LC is acceptable (even though not, in my opinion, correct) and only if it is based on the conviction that LC is a system intrinsically superior to DC (or whatever other systems are being abandoned--Cutter's Expansive Classification is still around in a few places waiting to be phased out, and there may even be a few American libraries using Brown's Subject Classification or Bliss's Bibliographic Classification (BC). And, though Richmond's cited address[6] is so oriented, such reasoning is probably not a large determinant in the minds of administrators, who tend to put administrative reasons before all else. By "administrative reasons" is meant cost differentials and associated factors such as proportion of staff time spent on non-public-contact operations, or the ever smaller proportion of the budget spent directly on the collection.

Thus, even the administrative reason in its purest form is not at all something we could judge to be in bad faith; that it is, nevertheless, incorrect, is the principal point of this essay.

But, whatever be our judgment on the good faith of administrative reasons, there are other professional reasons that must enter in. And it is in the very nature of any opposition between these two that the one which must win out is the professional, if we who are involved in the choice wish to retain claim to professionality--even those of us who are administrators.

Professionality is not necessarily a good quality. Another paper[10] describes both sides of this coin at more length, and the general point is that the idea of professionality derives from the verbal sense of professing, and that what the profess-or professes is something which he, in common with the community at large, confesses to be a value. This something, characterized by a value worth the dedication implied in the risk of professing,* is the logical intersection of inscriptions

---

*To profess is to take risk, to be willing to be sacrificed, whereas the confessor is (liturgically) one who is not a martyr.

and communication: <u>documentary information</u>. This is our pro-fession, that documen-
tary information is something of value, and that we are dedicated to this value: we
are its instruments. If we do not accomplish this communication* we have either
ignored or betrayed our profession. There are of course higher values, for which we
would be willing (at the risk of demonstrating to the sceptic that professionalism
is a bad quality rather than a good one) to set aside our professional values; we
could well be charged, otherwise, with a sort of fanaticism. But money as such is
surely <u>not</u> such a higher value; and purely administrative reasons, unless they appeal
to a <u>truly</u> higher value, are precisely such an anti-professional betrayal.

There is no such anti-professional betrayal involved in reclassification, to LC
or whatever, when the administrator does not see classification as an integral and
essential part of the value he professes. As Orne puts it, he has very little faith
in classification except "as a finding device"; when a patron wants information but
does not know whether there is a document which can communicate it to him, he can
discover its presence (or absence) "directly through the catalog, without using the
classification as a guide." Such an attitude clearly embodies an unawareness of the
extent to which "the catalog," namely the subject-headings in it, do not find needed
documents automatically, but <u>only insofar as they</u> (the subject-headings) <u>are an</u>
<u>adequate classification</u>, presented fragmentarily.[11] There is no reason why American
librarians should, in such overwhelming proportions, remain unaware that classifica-
tion is a total phenomenon, not one restricted to shelf-arrangement--unless we can
brand library education with a lack of commitment to its purpose of instilling
principles along with particulars.

What classification (in the sense of a total phenomenon) can and must accomplish,
if we are to uphold our commitment to the communication of documentary information,
will be the burden of §2, concentrating in §2.2 on refutation of the idea of subject-
headings as non-classificatory. What must be held in mind during the following de-
scription and analysis is that the conflict between the reasons characterized here
as "administrative" and "professional" must be resolved in favor of the latter, if
there is less than absolute proof that any classification can do as well as any other.
The burden of proof is thus upon the administrators, but the negative part of the
case will be presented anyway, in the hope of showing that the positive side could
not be argued successfully.

2.1 <u>The Qualities of a Classification Worth Changing To</u>

If one wishes, for professional reasons (generated, that is, by commitment to
furtherance of the professed value of documentary information), to re-classify the
collection of documents committed to one's care, one must look to the character-
istics of each classification that may bring about improvement--and of course do so
in comparison with whichever classification is now in use. What, then, are the
characteristics of any one classification such that it should be preferred?

At least on the idea plane, the comparison of extant classifications is largely
a matter of difference on collocation; but for all the argument over various colloca-
tions, and the general agreement that BC and Ranganathan's Colon Classification (CC)
are more elegant and logical in collocation than the other systems, this agreement
is not acompanied by any similar agreement about efficaciousness as a correlate to
this elegance. There is of course the further difference--one very similar to that

---

*As Ranganathan so well summarizes it: Every book its reader; Every reader his
book.

of collocation, but finer in detail--of what concepts are subordinated to what main classes. There is thus, on the idea plane, primarily the problem of order among concepts, a variation which can of course (as pointed out in "Documentary Relevance and Structural Hierarchy"[9]) cause failure or success in retrieval. And finally, more than any other aspect of conceptual order (or of any aspect of classification on any plane), the structure of the relationships which constitute the hierarchical subordination determines the ability of a classification system to respond well to the ineluctably central question of all reference service: if we have not found precisely what we need, what do we do next?

On the notational plane there is of course the general problem of hospitality, in array and in chain;[12] and in general the whole problem associated therewith, of the primacy of idea-order considerations over notational ones. Then there is the not wholly resolved dispute over the desirability of a structurally hierarchical ("expressive") notation.[13] There are theorists who see no possibility of a notation capable of satisfactory structurality. They may be right and they may have problem-examples in mind such as to defy all attempted solutions; but I doubt it, if we combine Ranganathan's sector (formerly "octave") notation[14] with my comma-device. It seems to me no less than self-evident that, if feasible, structurality is desirable, because only in this way do we achieve a ritual substitute for the intuitive decision as to "what shall we do next?"

The verbal plane, insofar as it is represented in systematic conceptual organizations (=classifications) by their indices, is not a central factor in the excellence of retrieval results; it is more an aid for the cataloger. Insofar as it is represented by the final products of alphabetical conceptual organizations (=subject-headings, thesauri) it is more debatable within the present frame of reference, since it is more clearly an influencing factor in retrieval--but this will be dealt with at greater length in §2.2 because of its peculiar hold on the imagination of the American library administrator.

What is the purpose of conceptual organization (whether systematic or alphabetical) in libraries? It is to provide a search strategy for documents relevant to needs, even when it is not known whether such documents exist within the corpus to be searched. But since, as enunciated above, the most fundamental question when names of documents are unknown is one that calls for something besides one-for-one matching, there must be the means for moving from the most specific concept(s) relevant to the need to those most nearly relevant, the next most specific, in a word. Ranganathan's notion of APUPA (Alien, Penumbral, Umbral, Penumbral, Alien) represents this desideratum diagrammatically; when we see that we must give up the hope for the desired specificity, we hope for the most relevant generality to be as close to our first point of attack as possible, and for there to be a gradual progression to the non-relevant, the "alien."

Two more factors enter in here, though; the first is that the APUPA principle cannot prevent the presence of relevances elsewhere ("distributed relatives"); the second is that specificity is not enough. Thought is correlation[15] and specificity is one kind of correlation, but surely far from all of correlation.[16] What is needed, then, is not mere specificity, but the ability to produce codes equivalent to any correlation implied by any document. This ability is here dubbed the concreteness of the classification, and I look upon it as the quality whose presence or absence is the most important of all in the evaluation of classification, largely because it represents the nexus of most of the crucial characteristics mentioned above.

The opposite characteristic is cross-classification, which occurs whenever only part of the document's conceptual correlation can be presented by the codes of the classification,*1 or when the correlates must be broken into several groups rather than concatenated into a single, concrete correlation. Every document which needs more than one classification code or one subject-heading (or one or more cf each) reveals the deficiencies of the classification in use.

We must be careful to understand the terms in the foregoing pronouncement. The "single code" does not of course mean a single digit, since by definition thought is correlation, and correlation implies multiplicity of correlates; besides, "every document" must be interpreted variably in terms of the principle of cotermineity, sometimes mapping macro-, sometimes micro-documents onto the conceptual organization, but in each case doing so at the level of the document being cataloged, and on the level of that document as a whole.

In light of the examination of the flaws of LC as a strategy for the Caribbean search we would hope for a classification that would make any occurrence of an idea easily recognizable by computer (or, for that matter, by patron), and likewise make each idea part of an expandable or contractable chain of subsumptions. Traditional enumerative classification tries to do these things, but does them in too simplified a manner, without preliminary analysis of complexes into elementary concepts. It is only with the advent of classifications that make use (both on the idea and the notational plane) of general categories that over-simplification can, even only hopefully, be eliminated. The seminal ideas for this development came from Melvil Dewey, but they did not reach fruition until the development of UDC, CC, and BC. And in light of the over-arching need for concreteness we could ask how LC (or, for that matter, DC) would handle correlations that go beyond mere "topic + place." To "art galleries--Havana" could well be added "administration" or "Italian renaissance painting and sculptures--restoration." LC's way out of this is to decide whether to classify under Italian renaissance art in general*2 or Italian paintings,*3 or Italian renaissance sculptures,*4 or restoration of works of art,*5 or Havana art galleries;*6 whichever terms have not been chosen for the (shelf-) classification can (perhaps) be rendered as subject-headings. But this device, which is defended

---

*1 It is almost ironic to see the welcoming smiles accorded to total non-correlation by those who so gladly accept unitermic indexing and the like; classification and what makes it valuable have not even come over their horizons, probably because they have had such poor examplars in view so long as to have been almost blinded by them.

*2 N6501-7413 is modern (= non-ancient) art by country; Italy, in the appropriate Table IV, is 411-423, divided the same as is Greece, 391-403, where renaissance (14th-16th cent.) =395; by analogy, then, Italian renaissance =415, which, when added to the base number, gives N6915.

*3 ND201-1113 is modern painting by country; we go through the same computation (with the same Table) to get ND615.

*4 NB201-113 is modern sculpture by country; for once there is a resemblance between analogous complexes: NB615.

*5 N8560.

*6 N910.H3, as explained in §1.3.

by some as the "multi-pronged" approach, is only an escape from the real need, namely
for an economical, intelligible, memorable, and (most of all, and in the senses ex-
plicated above) strategic and concrete code. LC cannot do it, as we have just seen;
UDC can. The complex heading "sculpture + painting of the renaissance in Italy--its
restoration in the art galleries of Havana" reads as [73+75]˙034(45)˙025˙:727˙7
(729˙11).

Such a number, it must be admitted, is long--longer than most DC or LC codes.
But it is short in comparison to the words necessary (as in subject-headings) to
represent the same concepts. It is longer, in fact, than would be the total number
of digits necessary to represent all its aspects in LC codes: 35 in UDC, 27 in LC
([NB615+ND615]+N8560+N910.H3), but this loss in economy must be balanced against
the gain in structurality of notation and memorability for search. Such a UDC code,
it must also be remembered, is not merely a number for shelf-arrangement, but is
also the basis for a classified arrangement of the catalog; indeed, it can be used
for this purpose alone, with the shelves arranged on some entirely different principle
(say, size for economy of space-utilization, or by broad classes and then directly by
author-code).

Finally, the UDC code, if used as the arranging system for the catalog, pro-
vides clerically recognizable permutation points. The cited number could also
create entries at [75+73]˙034... (painting and sculpture...), ˙034(45)˙025˙4:727˙7
(729˙11)*[73+75] (renaissance in Italy...), (45)˙025˙4:727˙7(729˙11)*[73+75]˙034
(Italy...), ˙025˙4:... (restoration...), 727˙7...(art galleries...), and finally
(729˙11)*[73+75] ... (Havana...). This last (as well as the entry for Italy), both
being place-indications and likely to be unsought, could have been omitted in the
printed or card catalog; but if the record had been stored electronically, all the
entries containing (729) and all extensions of it would have been easily recovered,
along with those at 972.9 and its extensions and correlations (in which last are
found historical and geographical treatments of the Caribbean area).

Although considerable emphasis is placed throughout this essay on the classified
catalog as an alternative to the alphabetical subject catalog, the presence of the
computer and the possibilities it opens up call for even greater emphasis on search
strategy for the consultation of non-conventional catalogs. Librarians must not
allow themselves to be branded as reactionary in moving from their previous attitudes
about construction and utilization of search strategies; nor can we afford to con-
tinue to take refuge in such excuses as are used to shield our subject-headings and
classifications from criticism, such as is being given here. It has been urged (in
conversation with Henry Dubester of the National Science Foundation's Office of
Science Information Service) that what is being attacked here is something LC subject-
headings admittedly cannot do[17] and that the attack is therefore not justified. Are
they above criticism just because they are modest in the face of new and greater
demands?

The computer offers a great deal; but before we can take fullest advantage of
it we must put our own house in order. We librarians must become more fully aware
of what our purposes are and of the theoretic bases of how we have been trying to
accomplish them. When (and if) this is achieved, it ceases to matter what someone
has said he does not expect his system to accomplish--except in the light of what
we see it _ought_ to accomplish or at least attempt to accomplish.

2.2 Various Fallacies About Classification in the Broad Sense

It is often said that "classification" is not really possible, and that since this is so we should rely on the catalog as the search mechanism. By this is meant the subject-headings in the catalog. But it has been shown that search strategy is based primarily on the question: "What shall we try now, our first try having failed?" Thus what makes a subject-heading catalog (alphabetical conceptual organization as against systematic) strategic is the structure implicit in it, by which we are led from Umbra to Penumbra. It was to show the possibility of an analysis of subject-headings into the form of an enumerative classification (and perhaps in some cases into that of a general-categoric one[18]) that the subject-headings of epistemology, ontology, and cosmology were diagrammed, as found on pp. 305-308 of "On Bibliography and Automation."[7] Such a diagram makes explicit and intelligible the step-wise organization of LC subject-headings. It maps out the whole of a strategy, rather than giving it to the searcher one step at a time.

To operate with a classification, even an enumerative one without general categories, is (compared to operating with an LC subject-heading catalog) like finding one's way across town with the aid of a map, as against asking directions at each street-corner. In this sense we can exclude subject-headings from the domain of classification; but we should not forget that classification in this strict sense of systematic conceptual organization, and subject-headings in that of alphabetical conceptual organization, are search strategies both--and are thus subject to the same criteria.

One argument against strenuous efforts to improve subject-headings, as well as against the substitution for them of a classified catalog, is that they are only appropriate to searches by the inexpert. That this argument is entirely fallacious can be seen from the degree of concern shown by a good many professional associations and governmental technical-information agencies for the development of thesauri-- which are merely more sophisticated versions of the subject-heading idea. Such thesauri are developed for the searching of groups of documents, in which searches it is postulated that the searchers, even though experts, do not know whether or not relevant documents will be produced as the result of the search. Thus the very development of thesauri belies the assertion that alphabetical conceptual organization is only worth the trouble in an environment of term-papers and how-to-do-it books. And the fact that it is experts who demand such developments also belies the associated fallacy that experts need no subject-catalog guidance--they know what they want, we are told, from citations and the like, and hence need only nominal-bibliographical catalogs.

Another fallacy is that of change from one classification to another, when neither is demonstrably inferior or superior, for the sake of monetary economy. That this is a fallacy can best be seen by the next step in the argument--or rather by the next step that is not in the argument: if one classification is as good or as bad as any other, this may be because they are both examples of "artificial" languages, and should both be avoided in favor of a "direct, natural" approach to bibliographical conceptualities, if such is available.

This is of course not obvious; otherwise, if there were advantage to be gained from HEA Title II-C, it would not accrue from the LC classification any more than from the DC--if neither is superior. But since there is supposed to be simultaneously available a direct and natural mode of strategic access, and if both other modes of

14

access are inferior because of their artificiality, it would seem that there is as much reason _for_ welcoming the subject-headings as there is _against_ change from one inadequate classification to another.

This argument is particularly potent when it is remembered that "classification" means, to most American librarians, only the shelf-arrangement, and hence that re-classification implies the changing of document-labels, circulation cards, circulation-card-pockets, etc.; whereas change from one search strategy (as catalog arrangement) to another, while it surely requires some physical alterations, can at least be accomplished without changing the shelf-arrangement. However, we cannot really be sure that leaving the shelves untouched in the course of search-strategy change would be wise, at least in public libraries, since it has been found that a far higher proportion of their patrons expect the shelf-arrangement to assist them than do those of other types of libraries who seek assistance through the catalog. (The unpublished research data on which this conclusion is based is from a project conducted by Dr. M. L. Bundy for the state of Maryland, in 1966. The number of persons surveyed was 21,138; 33 percent used the shelf-arrangement for searching and 14 percent the catalog. The remaining 53 percent used the periodical indices, had discussions with librarians, used record players, etc.; it is significant that the 33 percent was the highest single type of use, with 17 percent for use of reference books next highest.)

Working back through these arguments, then, it seems that subject-headings provide an inferior search strategy--were this not so it would be the more efficient provision of them which would represent the main advantage provided by HEA. But the fact that subject-headings are inferior is demonstrated by the current development of essentially similar strategies, in the form of technical thesauri; and this development simultaneously demonstrates the need for the improvement of search strategy for the sake of expert and inexpert alike. But that the "natural" search strategy is in fact superior to the "artificial" is only another fallacy of the same sort, because it is subject to exactly the same criteria as those by which we can judge classifications (in the strict or popular sense). And such criteria cast a light which shows that, since the fundamental problem of retrieval is "What, then, next?" and since this problem requires a wide-range strategy, alphabetical conceptual organization is inferior to systematic just insofar as it does not offer as good a solution to the fundamental problem.

2.3 "A Place for Everything is Good Enough"

An opinion at least apparently shared by Richmond and Orne, and surely by a great many others as well, is that systematic perfection is a less important characteristic of a search strategy than the ad hoc provision of a place for everything. In a sense that is true--in the sense that we store away our own collection, of documents or whatever else, in an order and in locations which in no way reflect their meaningful contents or usefulness. But this situation obtains only up to the capacity of the memory of the individual; when that limit is exceeded, surrogates must be created to assist the memory. And when the "memory" that is being thus supplemented is shared by all those who are cataloging documents by thousands each year, it is necessary that there be more than merely ad hoc ways of getting back to these documents upon demand. Thus are developed both codes of author/title cataloging

and systems of conceptual search strategy. Both such sets of rules are designed, not for analytical bibliography with its concern for the book--and even for the individual copy--but for systematic bibliography with its concern for the work and for the relations between works within the corpus. Author/title codes and classification systems provide the systematic foresight that prevents the one-by-one incorporation of documents from forming not an organic corpus but a disorganized mass.

An example of conflict in citation order and predictability can be helpful here; the point of departure will be DC (since, while I am horrified at the idea of re-classification to LC, I am not too happy about the permanent retention of DC either). Our test document is an Australian union catalog of scientific periodicals, arranged by subject. We recognize Bibliography at once as the main class, and the most obvious order of analysis of terms runs: Bibliography--Catalogs--Union--Science--Periodicals--Australia. When we look for a beginning-point in the index we find 017, and since it is a catalog of libraries as such, we extend this to 017˙1, "union catalogs of public subject-arranged libraries." And we see that the topic "scientific periodicals" cannot be attached to 017˙1; however, we can at least add on "Australia," giving 017˙10994, but the central point, the topics covered in the collections indicated, cannot be shown. So instead of starting from "union catalogs" we can start from "subject bibliographies," 016, to which can be added "scientific" 016˙5, but not "periodicals and Australia," which would give 016˙5050994, since 016˙505 would be read back as "a periodical on the bibliographies of science." Since two of the first duties of an information language are explication of homonyms and consolidation of synonyms--namely, in a word, the elimination of ambiguity--no such misreading of the original analysis can be allowed. (Such a misreading might be defended by those who propose that classification is no more than pigeon-holes--places for everything, but nothing more than places-- in terms like: "Well, when or if someone finds this document, he'll understand what the code really means, and until then, it gives us a unique place to keep it."

So we can go only to 017˙10994 or to 016˙5, neither of which represents the full concreteness of intersecting categories which the document defines. But there is a way, if our citation-order analysis is altered to Science--Periodicals--Bibliographies--Australia. Then a code can be constructed to accomodate all present categories: 505˙0160994. But, assuming that we have a policy that dictates placement of all bibliographical items together, such a solution results in radical unpredictability, and thus cannot be tolerated.

What can be done with this problem in LC? First of all, of course, we do not even have the option of moving it out of Z into Q; and in Z we can code it as Z7403 (Bibliography--Science--Periodicals) only, certainly not as Z695.83 (Library science--Union catalogs); for a classification whose "specificity" is so much vaunted, this is highly un-concrete. We might choose Z975 (Library science--Catalogs--Australia) but that too is too un-concrete. With LC subject-headings we can have "Catalogs, Union," but it cannot be extended for greater concreteness. A better solution (not led to from "Catalogs, Union" by sa, though) would probably be "Periodicals--Bibliography--Union lists," which can be divided by place, unlike the first heading, giving "Periodicals--Bibliography--Union lists--Australia." But there is still no mention of Science. To get this essential point in we must have another heading: "Science--Periodicals--Bibliography." (It is interesting to note the distance that these headings lie from the beau ideal of alphabetico-direct theory.)

Now, before we go on to coding this in UDC, we must examine why there is restriction in DC on going all the way when we begin with 017·1 or 016. The third duty of an information language (after the first two given above) is to establish rules of formation: a systematic conceptual organization is an artificial <u>language</u>, and must therefore have both a vocabulary and a <u>syntax</u>. But there are no <u>rules of</u> formation (syntax) in DC, except the implicit <u>one that</u> can be stated algebraically thus: (A), ((A)B), (((A)B)C), ((((A)B)C)D) etc., which can be interpreted that each last term is a modification of the complex that precedes it. This situation obtains in DC largely because of the original (and still obeyed) desired to have a <u>pure</u> notation, which desire prevents the use of any overtly relational or syntactical codes.

In UDC, the various aspects are similar to those of DC except in the use of punctuational symbols in place of DC's purely numerical facet indicators. Catalogs-- Subject arranged--Public collections--Union, is coded as 017·11; Australia is (94); Science is 5; Periodicals is (05) or :05. (This last distinction is an advance over DC, allowing as it does a choice between "periodical" as a form and "periodicals" as a form-topic. If such a distinction had been available in DC, it would seem that we could have gone beyond 016·5, not to 016·505, which would necessarily involve us in an ambiguity, but to 016·5:05. But even this would be subject to an ambiguous misinterpretation as "[the topic of] Periodicals dealing with the bibliography of science." The only escape from this trap is the use of square brackets, as exemplified below and fully discussed elsewhere.[19]

If we utilize the inverse of the filing order of the UDC symbols,[20] we would find the basic subject, put place (...) after it, then outer-form (0...); but we have here not one but two basic subjects, 017·11 and 5. (Note that "union catalogs" or "subject bibliography" cannot be used as terminal subdivisions, as in DC.) These two basic subjects (as is the case with any multiplicity of basic subjects drawn from various parts of the schedules, and treated intersectedly in the document at hand) can be joined by the symbol : ,which is interpreted as "logical intersection," giving 017·11:5 for "union catalogs . . .in science." We have seen from our attempt to do all this in DC that "Australia" must modify "union catalogs," not "scientific periodicals"; so we can modify as needed, 017·11(94), and then intersect this with the other basic subject, giving 017·11(94): 5. Now we wish to add the further concreteness, namely that the content of this Australian union catalog is not merely "science," but "scientific periodicals"; so we try to add :05 to 5, giving 017·11 (94):5:05 But this formulation is also subject to a misinterpretation, namely as "periodical (as a topic) <u>on</u> Australian union catalogs in science." But there is a solution. Instead of stopping us, like DC, with ((A)B), the : has allowed us to go on to (((A)B)C)*; but we still cannot simply add D at the end, if we wish to avoid the given misinterpretation. Instead, we can explicitly use algebraic sub-groupers (square brackets)[21] thus: 017·11(94):[5:05], which is interpreted so that only 5 is modified by :05, and the whole complex reads back "union catalogs, Australian, on science periodicals"; algebraically, then, (((A)B) [(C)D]).

We can far more nearly exactly predict the main classified entry for such a document in UDC than in either DC, LC, or (though this is added only hypothetically)

---

*Note that there has been a change in the semantic filling of each algebraic symbol, as against the (ambiguous) DC order 505·0160994.

LC subject-headings. In DC and LC, one of several only partially concrete codes can be chosen, and it is therefore impossible to predict which will be used; this is a sort of cross-classification, which is thus seen as not merely an abstract horror in the minds of classificationists but as a horror in use as well.

2.4. Some Problems Relative to UDC Filing Order

One aspect of the use of UDC that comes under fire most often (though erroneously) is connected with the problem of predictability--namely, the filing order between UDC codes. As against a DC code with its absolutely pure notation (which only appears to be mixed because of the presence of the wholly content-empty decimal point), a UDC code utilizes a mixed notation which makes possible the easy indication and interpretation of the various facets represented within it. This factor should indeed make the filing and locating of documents and/or surrogates far easier. And there is certainly no virtue in the filing order of LC codes, even though it is programmable-- since they only locate, as if we were back before the days of relative classification, i.e. back when a shelf-code was just that, an assignment of a particular document to a particular place on a particular shelf in a particular cabinet in a particular room. What is gained by the mixed notation in UDC is the ability to file by "empty," "partially empty," and "full" facets.

The DC code 894˙51110409003 (20th century Magyar lyrical poetry; see §3.21) places its document in order after that coded 894˙51110409002 (894˙51110409004 does not occur, since there has as yet been no such period of time); its facets can be shown by spacing-out, thus: 894˙511 104 09003. But a code for "20th centry Magyar poetry," which, being obviously more general, must come before in the file, cannot do so except by facet-by-facet comparison (i.e., partially empty compared with full): $\frac{894.511 \quad 1 \quad 09003}{894.511 \quad 104 \quad 09003}$; but the fact is that such a preliminary operation is not normally performed nor even recognized as necessary, so that the more general document comes later: $\frac{894.51110409003}{894.511109003}$, because of the accidental and misleading comparison between the ninth digits of each code. With UDC codes for the same concepts, on the other hand, the presence of the facets is entirely clear: 894˙511-1"19" cannot come anywhere but before 894˙511-14"19", because -1 (partially empty) comes necessarily before -14 (full). This phenomenon can be referred to analogously: DC filing is "letter-by-letter," UDC "word-by-word."

The "presence" of wholly "empty" facets too can be a problem in a classification using either a pure notation or a mixed notation without facet indicators. An analogous case here is the filing of main entries for Biblical documents: "Bible. New Testament. Greek . . ."comes after "Bible. English . . ." not at all because "N" comes after "E," but rather because whole comes before part; this can be realized quite easily if we just imagine where "Bible. Swahili . . ." comes in relation to the two mentioned entries, namely between them. Since classification too is a species of systematic bibliography, it too must seek such groupings, and is assumed to be able to do so more easily because it uses a notation and an index, whereas author/title cataloging must rely on the natural language and self-interpretable terminology. But what happens constantly in DC is that ritual interpretation is impossible because of the need for levels of generality, which forces the use of multiple introductory zeros; thus at 942, the first enumerated sub-heading is for the most ancient historical period, 942˙01, forcing the standard subdivisions to

adopt an extra zero: 942˚001/˚009. And it is not impossible to find that even the double-zero is preëmpted for more concrete tasks, so that the standard subdivisions require three zeros.

Other charges against the filing order of UDC complex codes, especially when formed with the colon-device, are brought by Metcalfe,[22] but his examples of chaos seem as lucid as could be hoped for to me (though not his use of the ˚00 auxiliaries as intermediaries). What makes his example at least apparently chaotic in its file-order is the very paucity of coloned-on numbers which Metcalfe comes up with; certainly if he had used a base number like 016, to which very nearly every other number may be added, he would have had no objections to make. What it comes down to in the end is that he rejects non-alphabetical collocation, and does so all the more vehemently when it is represented by only a partial selection of such coloned-on secondary numbers. It is even implied that if, instead of retaining these borrowed secondary numbers, sub-classes were directly enumerated, all would be well.

The easy interpretation of the explicit complexity of UDC codes compares very favorably indeed, in my opinion, to the tricky interpretation of the only implicit complexity of DC codes; this, together with the generally shorter quality of the former, makes one wonder what constitute filing problems to those who find fault with UDC order. There follows a comparison of order between UDC codes and DC codes for the same conceptual complexes (it will be noted that DC is often either un-concrete or unspecific).

```
894˚5 [Finno-Ugrian literature]────────→894˚5
894˚5"19" [20th century]──────────────894˚509003
894˚5˚085 [its oral presentation]         [oral presentation is not specifiable]
894˚5˚085"19"
894˚5-1 [poetry]                          [forms of the family of literatures
894˚5-1"19"                               are not specifiable, since they would
894˚5-1˚085                               take the form 894˚51, which is already
894˚5-1˚085"19"                           enumerated for Ugrian literatures]
894˚5-14 [lyrical]
894˚5-14"19"
894˚5-14˚085
894˚5-14˚085"19"
894˚511 [Magyar]──────────────────────894˚511
894˚511"19"───────────────────────────894˚51109003
894˚511˚085──
894˚511˚085"19"─
894˚511-1 ─────────────────────────────894˚5111
894˚511-1"19"─────────────────────────894˚511104
894˚511-1˚085──────────────────────────894˚51110409003
894˚511-1˚085"19"────────────────────894˚511109003
894˚511-14─
894˚511-14"19"─
894˚511-14˚085─
894˚511-14˚085"19"─
894˚511-2────────────────────────────894˚5112
894˚511-2"19"──────────────────────────894˚511209003
894˚511-2˚085─
894˚511-2˚085"19"─
894˚511-2-14─
894˚511-2-14"19"─
894˚511-2-14˚085─
894˚511-2-14˚085"19"─
894˚541 [Finnish (Suomi)]──────────────894˚541
```

3.11 A Counter-Proposal: Centralized Classification by UDC

LC has a great many more enumerated classes than does any other available general classification, in particular DC. DC has fewer than UDC or BC as well. CC probably has fewer than DC, though. So, conclude the naive, LC must be the most specific, and UDC next so. This would be like believing that a language with $2x$ words in its total vocabulary, but with no formations possible but those of the pattern $x_i + x_j$, has twice the expression-possibilities of another language with only x words but with formations of the patterns $x_i + x_j$, $x_i X x_j$, x_i / x_j, $x_i - x_j$, $x_i + x_k$, $x_i X x_k$, x_i / x_k, and $x_i - x_k$ possible. Thus it is the capacity for precision, as defined above as the ability to produce codes equivalent to any correlation implied by any document, which is the most powerful factor in the expansion of the total number of expressions possible in any language, natural or artificial;[23] with its powers of expression in levels and rounds, CC might well have a larger gamut of developed expressions than any other information language, despite the smallness of its list of elementary terms. In any case, there is no reason, when we add syntax to vocabulary in our computations, to accept LC as more specific than UDC even though it may well be far more so than DC.

But there is need for more than additional specificity; there is need for more than just a stop-gap solution, more than just monetary economy (especially if the main advantage rests in a different sort of change than that contemplated). This problem is not one that needs to be solved just for the sake of incipiently large collections; it needs solution for that of already large collections, in order for them to be used to their fullest. It is often proposed that the smaller the collection, the more thoroughly it must be strategized for sufficiently numerous relevances to be produced when the need arises. This is nothing if not trivial; the larger the collection, the _more_ the need for a thoroughly developed search strategy.

What is proposed, then, is that UDC be the classification to which libraries change when their seams begin to burst, whether they are now using DC or even LC. For those using DC now, the change can be relatively painless (see§3.21); there would be at least the possibility of intercalation of new and old, instead of the usual technique of creation of two side-by-side collections, differently arranged. With collections now arranged by LC, of course, intercalation would not be possible-- but service-improvements would be.

This is only an abstract proposal, at least up to this point; but the problem is not abstract, it is very real. What makes the new situation so new (with HEA Title II-C) is that classification-effort is reduced by its centralized provision. So, what can make UDC re-classification similarly attractive may well need to be the same sort of actuality; what may be needed is a return to a function proposed for itself by the Institut Internationale de Bibliographie, the ancestor of the present Fédération Internationale de Documentation: the function of centralized classification of the research-journal articles of the world. (Attendant difficulties will be outlined in §3.4.)

This is a tall order, even when one's hope is only for centralized classification of books, not for journal articles. But as a stage just below that of centralized UDC classification, I propose change to UDC classification done by each institution for its own purposes. The fact that this, despite the lack of the economies of centralization, can produce the desired results, is largely due to the theoretical superiorities of the UDC over both DC and LC; problems of implementation will be summarily discussed in §3.3.

20

3.12 Why Not a Proposal in Favor of CC or BC?

Since what is needed for the improvement of search-strategization is a change from enumerative (LC) and semi-enumerative (DC) to general-categoric classification, it may well be asked why CC is not being advocated, since it is the most thorough-going exemplar of the general-categoric approach. However, both it and UDC (but not BC) are rigorously criticized in regard to their categoric structures in de Grolier's excellent Study of General Categories.[24] The fact, though, that neither CC nor UDC is without flaw (as de Grolier surely shows), and that their flaws are differential, is not what leads to a preference for UDC. The major reason for not advocating CC or BC is their notations; not that they do not do what is expected of them (namely, primarily, that they "mechanize" the order--in array and in chain--of concepts), but that they are so alien to what we expect a library notation to look like, that it would be very surprising if they could be widely acceptable in American libraries. An additional advantage of UDC is its close resemblance to DC; a policy of "osmosis"[25] could be adopted in re-classification from DC to UDC which, while it would be less than perfect in having on the same shelf documents collocated somewhat differently, could at least allow intercalation--as against the necessity of parallel and separate collections when re-classification from DC is to LC or CC.

It might be argued that there is a fairly strong resemblance between LC and BC, and that the same advantages would accrue to a change along that axis as would be the case for DC/UDC change. The resemblance is there, surely, but it is almost entirely a formal one, depending on the fact that both use a mixed notation beginning with capital Roman, then Arabic numerals, etc. But the collocation of main classes is far less similar than with DC/UDC: only class Z is anywhere close to similarity in content.

3.21 A Brief Comparison of UDC and DC

UDC is a classification governed by the FID: it is modeled on the DC, but considerably modified both in collocation and synthetic structure; its notation is mixed (consisting of the same sort of numericals found in DC, plus punctuationals; generally without alphabeticals; decimal; highly flexible). Many of these points can be seen from examples given above and below.

But UDC varies from DC in other things. It is without as pronounced a Western slant, since the intention was to be universal. The United States likewise is given less predominance than in DC. The one class that is most fully developed (in terms of enumerated specifics) is 6, "Applied Sciences. Medicine. Technology." In the abridged English edition, class 0 occupies three pages, 1 three, 2 three, 3 twenty-two, 4 (currently being vacated and transferred to 8) two, 5 twenty, 6 fifty-two, 7 eight, 8 one, 9 two, and the general auxiliaries thirteen. There is great need, obviously, for further development in the humanistic and (to a lesser extent) the social-scientific classes, but even the sparse enumeration here listed is very greatly expandable by the synthetic-formation possiblities. The 78 schedule (music) occupies only one page, yet it is capable, by internal combination, of representing such a concrete topic as "performance practice of baroque liturgical organ music," giving 783'1: 786'6'091'034'7; DC can only come close, with 786'60932 (Organ--History--Baroque) or 783'073 (Sacred music--Performance); LC is no better, giving ML554 (Organ--History--Baroque), ML604 (Organ--Performance--Baroque), or BX9187 (only applicable if the document deals with instrumental music--including the organ--in the Presbyterian church; why there is no general class no one can guess, nor any for

other church bodies); LC subject-headings give "Organ--History," "Music, Baroque," "Music--Performance," and "Church music--History and criticism," some of which would be likely to be excluded for the sake of economy.

The largest present body of utilizers of UDC[26] consists of European special libraries, primarily in the scientific and technical areas; the examples used here should show, though, that there is no reason to fear that it is incapable, because of less enumerative specificity, of handling humanistic subjects well too.

The revision of UDC is constant, and is vested in committees of subject- and classification-experts (some of very broad responsibility like that dealing with 1/2, others narrow, like that for 621·3). There is no likelihood that either retention of DC or change to LC will gain the concerned library anything at all in greater currency of inclusions or of terminology.

The codes that are created by UDC are often longer than those created by DC-- for the same document, of course. However, the DC code is not shorter because it represents the same concreteness more compactly; it is shorter <u>only</u> because of its lack of concreteness.

DC is also quite clumsy to interpret because of its reliance upon pure-numerical notation and fractioning by threes; for a code like 505·0160994 this gives a book-spine code that looks like $\frac{505}{.016}$ or even $\frac{505}{.01}$, whereas it ought to be $\begin{smallmatrix}505\\.016\\0994\end{smallmatrix}$, or (as in UDC) $\begin{smallmatrix}5\\:05\\:017\\.11\\(94)\end{smallmatrix}$ $\begin{smallmatrix}099\\4\end{smallmatrix}$ $\begin{smallmatrix}609\\94\end{smallmatrix}$

The UDC code has been re-arranged to begin with the same element that the DC code <u>must</u> begin with if it is to succeed in being concrete. Note that the UDC advantage is that it can be precise within the limitations of the policy of main (classified) entry under 0, while DC cannot. Part of the difficulty in interpreting a code like 505·0160994* is in knowing <u>where</u> each new facet begins, and in knowing <u>what</u> it means. In the given case a single 0 begins each facet, but if there were a different semantic filling demanding a change in order to Science--Periodicals--Australia--Bibliography, the 016 could not be retained in that form, since it would have to be interpreted as a time-qualifier of 994 (="early period"); it would need to become 0016, giving 505·09940016. A UDC code, on the contrary, is unalterable in semantic filling.

Another example of the difficulty of interpretation (by a reference librarian or a patron) of a DC code which succeeds in representing the whole correlation with concreteness is "20th century Magyar lyrics," 894·51110409003(§2.4). The multiplicity of 0's and the clump of 1's would make this far harder to interpret than the same correlation expressed in UDC--894·511-14"19". The conclusion (suggested in conversation by A. J. Wells, editor of the British National Bibliography) may well be

---

*We must guard against comparing UDC 5:05:017.11(94) with $\begin{smallmatrix}505\\.016\\0994\end{smallmatrix}$, since that is not the way DC is used; we must compare current recommended practices in each system. (The use of virgules to separate facets has been suggested at the LC-DC Office; this would give 505/·016/0994 or the like.)

that UDC, with its clear facets and easy permutability, is so much better for catalog-arrangement as to make consideration of DC for that function pointless, whereas DC (at least in its pre-17th edition form) might well be better for shelf-arrangement, where concreteness (in my sense) is not as useful. My personal response to this suggestion would be partly favorable, except in regard to the variation between the two collocations; it would seem more logical to arrange the catalog by the full UDC code, permuted, the shelf by the first or first two facets.

The primary (and original) function of UDC was for the ordering of surrogates, not of documents; indeed this was once the case with DC. The presence of clear facet indicators, though, makes UDC (and CC too, of course) ideal as a catalog-classifier. An example of how this could work is given at the end of §2.1; such a device eliminates subject-headings with their total unpredictability and ignorance of the APUPA desideratum. And, since it makes the catalog itself far more effectively browsable, it frees the shelves from the necessity of being a search strategy on their own, and brings the surrogate- and the document-arrangement under the same rubric; the catalog then simply provides all the multiple (main + added) access vistas that the shelf-arrangement obviously cannot, and all is done in terms of one and the same set of principles.*

3.22 A Brief Comparison of UDC and LC

We have already discussed the virtual impossiblity of computerized search of an LC file (§1.3). Here we will take up some of the logistical difficulties involved in preference of UDC over LC. There is, more than anything else, the difficulty of obtaining a complete full English UDC.[+] There is reason to hope, though, that the current research project on UDC and mechanized searching[27]--especially in its preliminary phase of collation of a current full English edition on magnetic tape--will provide the copyright-holder, the British Standards Institution, with the basis for a complete full edition in main class volumes, perhaps within the coming year.[28] When this occurs there will not need to be as much concern over the apparent "absolute" size-superiority of LC over UDC--though §3.11 has shown how little such a comparison means.

That LC is unprogrammable is not, and cannot be, proven--and I shall make no such attempt. What is clear enough, given the description of search strategy provided above, is the relative ease of doing so for CC and for UDC. An experimental program on the seismological section of UDC has been reported on by Caless;[29] it is no surprise at all that he concludes that electronic-tape searching is feasible in actuality as well as in theory.

---

*It is one of the strangest developments in American librarianship that we have made so little attempt to unify the strategy of shelf and of catalog. DC and Sears, LC and LC subject-headings, DC and LC subject-headings--none of these pairings has anything to recommend it except simultaneous availability; LC and Sears would make every bit as much sense.

[+]There are full schedules for some classes and sub-classes, but not for all 0/9; hence what can be obtained is full but not complete. On the other hand, there is not too much difficulty in obtaining complete abridged schedules--these might do for small-to-medium-sized libraries, but few of them are under pressure to re-classify--at least prior to HEA.

It is thus definitely shown that "coordinate" or "post-coordinate" (a better term would be "unarticulated," or "discrete") indexing, with its denial of all correlational value, is not by any means the only search strategy adapted to sophisticated mechanized use. Classification--hierarchical, "artificial," intellectual, and all the rest--is henceforth to be excluded from the progress towards "information retrieval" only insofar as it is inflexible, unstrategic, and inhospitable--such, in other words, as has been described above and in "On Bibliography and Automation."7 And the capital instance of all these defects, as clearly as anyone might ever hope or fear, is LC.

3.3 What if Centralized UDC-Classification Cannot Come About?

There are various solutions to the problem of classification and re-classification, and they can, in many cases, be stated in sets of bifurcations or dichotomies. We have the dichotomy Classified Shelves/Unclassified Shelves; the bifurcations Alphabetical Catalog/Systematic Catalog, Institutional Surrogation/Centralized Surrogation, Card Catalog/Book Catalog, Dictionary Catalog/Divided Catalog, etc. While there is a tendency for some of the limbs of pairs of these choice-situations to become rigidly associated, there are more such popularly associated pairs than is logically necessary. For instance, a library with classified shelves is expected to have an alphabetical catalog; one with a book(-form) catalog is expected to have a divided catalog. The situation at hand, though, is to decide what goes along (logically, though not necessarily popularly) with the choice of Institutional as against Centralized Surrogation.

My observation of classification students at the School of Library and Information Services, University of Maryland, as well as my own remembrances from working with DC (at Milwaukee Public Library) and with LC (at Florida Atlantic University) would lead to the conclusion that, in the absence of any outside (centralized or cooperative) assistance in conceptual bibliography, the use of UDC is both more economical of time and energy, and more consistent in results over time and/or among multiple personnel.

It is more economical of time (a) in that its index, while not perfect, is far superior to that of the seventeenth edition of DC; and the more compact layout of the schedules enables a more rapid survey of the generic/specific situation. The same is also true (b) because of the faceting and the unambiguous notation, which conspire to prevent the need for going from place to place seeking solutions by analogy.

To obtain the DC code 894.51110409003 (§§2.4,3.21) it is necessary to go from p. 1848 to p. 1728 to p. 472 to p. 1140 (without any explicit lead to this last); we are now at the first class code 894°511; then to p. 1120, which tentatively lets us add the next facet 104, giving 894°511104, though to verify it we go on to p. 1114 (note under 811°02-°08), to p. 1104, none of which changes our original tentative decision except to make us wonder why "lyric" is 821°04 on p. 1120 and 811°04 on p. 1114, but 808°814 on p. 1104; then back to p. 1148 to add the last facet "20th century," giving the whole code 894°51110409003. By contrast, in UDC we go from p. 203 to p. 192 to p. 143 (main class) to p. 142 (literary form) to p. 22 (period), accumulating 894°511, -14 and "19" which together give the whole code (better anyway, as well as easier) 894°511-14"19".

It is _more_ _economical_ _of_ _energy_ in that its flexibility prevents the frequent occurrence of cross-classification by arbitrary choice of a facet to be excluded. It thus prevents frustration, the greatest energy-drain on the classifier using either LC or DC (_see_ examples in §§2.1, 2.3, and 3.21).

It is _more_ _consistent_ _over_ _time_ _and/or_ _among_ _multiple_ _personnel_ because its use can be taught (as with CC as well) in terms of guiding principles, rather than by gradual accretion of particular problem-cases and of their ad hoc solutions--a technique better fitted to in-service training anyway. In particular, the availability of a notation that allows an undiluted citation order (the example of the union catalog in §§2.3 and 3.21 is sufficient here) leads to such consistency.

It must be admitted that high flexibility can lead to bizarre results--but only if the classifier has no grasp of the fundamental principles of search strategy. If he does not, he may produce such a code as $017^\cdot1(94):5:05$, i.e. without the necessary sub-grouping square brackets at the end; or he might assume that the same square brackets are needed even when the citation-order policy calls for 5:05..., which is not the case (though it is not positively harmful); or he might come up with an order than seriously distorts the correlation of the intended semantic filling, such as $994:05:5:017^\cdot1$.

3.4. How Can Centralized UDC-Classification Come About?

The set of advantages given in §3.3 is better, by quite a distance, than nothing, but it cannot compare with the advantages of work once done and in need of no meliorative repetition. How is this desideratum to be encompassed?

One possiblity, suggested in §3.11, would be for the FID to undertake to return to its original function of a central document-classification-provision agency. But it is highly doubtful that FID will be able to do more than provide moral support, plus (perhaps) consultative services. In any case, it is the coupling of classification with author/title and descriptive cataloging centralization that makes LC and LC subject-headings such an administratively attractive solution. Who, then, will be the "sub-contractor" in this enterprise?

There seem to be two ways most likely to lead to a profitable conclusion. The first is quite unrevolutionary; it would be for a cooperative network of libraries to undertake, between them, to create a consolidated set of surrogates classified by UDC. If this set of surrogates were stored on magnetic tape, the diversity between the various collections would create a totality which might well be marketable and easily distributable. An electronic network arrangement could also eliminate the simultaneous classification of the same document by the several cooperating libraries. Perhaps the first step toward such a network would be the demonstration, in a manageably middle-sized _general_ collection, preferably already automated, of the high utility of such an effort. (It would be most advisable, of course, for there to be outside funding for such a public-spirited experimental effort). Once this is accomplished, the chances for the establishment of the network might well be stronger.

The second way is more revolutionary; let the Library of Congress establish an office for centralized UDC classification in cooperation with FID, (or possibly with the British National Bibliography), thereby causing Title II-C of HEA to provide three sorts of search-strategic information instead of two. For the library itself,

then, or for libraries immovably attached to LC and the dictionary catalog, the LC classification codes and subject-headings will continue their traditional, if disputable, service. For those using DC now but aware of the imperfections of it (and of Sears or LC subject-headings used in tandem with it) as a search strategy, there would be opened up the wholly new prospect of a differently arranged catalog, which either on magnetic tape or on cards would provide throughgoing correlation, clerical permutability (multiple access), and the APUPA principle (for browsing)-- and they may continue to use DC (as suggested by Wells; see §3.21) for shelf-arrangement.

Or, to go one more logical step, the UDC Office might well (even if not at once) entirely supplant the DC Office, since the differences in collocation between the two, and the resultant disparities between the classified-catalog code and the shelf-arranging code, might prove more disturbing to the patron than a more thorough divergence.

Such a wider (and more easily acceptable) divergence would emerge if UDC were used to organize a classified catalog of documents arranged on the shelf by LC. This solution would not be acceptable, however, because the Bundy findings (reported in §2.2) cannot but prevent us from acting as if shelf-order without any symbolization of semantic filling (as in LC, where the semantic filling of the codes is entirely opaque without the appropriate schedule to refer to for each and every code) is good enough for our patrons, who "don't know what they want anyway, so it really doesn't matter."

An example of the divergence in collocation between DC and UDC would be the codes for "history of philosophy," which in DC (and thus on the shelves) occupy 180 through 199, in UDC only 19 through 199, thus clearing 18 for "philosophy of beauty" (although it is not too effectively used for that).

4.0 Epilogue

Is it too late? Many libraries have already changed from DC to LC, and are unlikely to be willing to spend the additional funds necessary to rectify the mistake (hopefully, it has been shown here to be such).

But the availability of centralized cataloging, if accompanied by UDC classification, might rectify even this desperate situation, since it would eventually produce the basis for a change in the catalog--which after all, not the shelves, is the crucial nexus for search strategy, since shelf-arrangement cannot help but be dominated by the problem of distributed relatives and that of the physical unavailability; it is the multiple access vistas of the catalog, together with adequate correlation (concreteness) APUPA, and notational structurality, which solve the problem of classification--or, more broadly, of search strategy--namely, "What, then, next?"[30]

REFERENCES

1. See e.g., the letter by L.B. Archer, "Ultimate Conversion," Library Journal, 91:4868, Oct. 15, 1966, which reveals the extent to which dissatisfaction with the 17th edition of the Dewey Decimal Classification (DC) creates pressure for a solution--and also shows the extent to which the only solution that is even considered is change to LC.

2. Richmond, Phyllis A. "General Advantages and Disadvantages of Using the LC System." In Proceedings, ALA Pre-Conference Institute on LC Classification. Summer 1966, (to be published by ALA, 1967).

3. That she is also entirely aware of disadvantages--little as she concentrates on them in the cited address--is quite evident from her letter: Phyllis A. Richmond. "Switch Without Deliberation," Library Journal, 91:4870, Oct. 15, 1966.

4. So much so that members of the Association of Research Libraries, each with collections of 1,000,000 titles or over, often catalog from 40 percent to 60 percent of their current acquisitions for themselves; cf., e.g.,Ellsworth, Ralph E. "Another Chance for Centralized Cataloging," Library Journal, 89:3104-3107, Sept. 1, 1964.

5. Title II, Part C, is the relevant section, as described by John W. Cronin, "The Library of Congress National Program for Acquisitions and Cataloging," Libri: 16:113-117, 1966.

6. Richmond, op. cit.

7. For a discussion and validation of these three elements of cataloging, see the set of papers by the present writer, "On Bibliography and Automation; or, How to Reinvent the Catalog," Libri, 15:287-339, esp. 298-299, 1965.

8. We leave aside the consideration of the computer's printing of cards to be filed manually in an otherwise conventional card catalog, as advocated by Frederick Kilgour at Yale Medical Library, "Development of Computerization of Card Catalogs in Medical and Scientific Libraries." In Herbert Goldhor, ed., Proceedings of the 1964 Clinic on Library Applications of Data Processing. Urbana, Ill., University of Illinois Graduate School of Library Science, 1965, pp. 25-35.

9. It should also be noted that those codes which represent concepts of which the Caribbean area form, in extension, a part of the first order, should also be predictably relevant. A convention which would enforce such a predictability is described in my papers "Documentary Relevance and Structural Hierarchy," Information Storage and Retrieval, 3:13-18, Aug. 1966; "Coterminous or Specific; a Rejoinder to Headings and Canons," Journal of Documentation, 22:319-328, Dec. 1966, but it is not rigidly enough conformed to in the practice of cataloging at the Library of Congress to have been considered as a potentially fruitful angle of attack.

10. "What It Is to Be 'Professional'," (Iconoclast, in press).

11. Cf., e.g., Richmond, Phyllis A. "Cats: an Example of Concealed Classification in Subject Headings," Library Resources & Technical Services, 3:102-112, Spring 1959.

12. The latter of these is adequately solved by the radix-fraction principle discovered by Dewey, but the former is not really solved in any available classification ; cf. my paper "A New Device for Achieving Hospitality in Array," American Documentation, 16:245-246, July 1965.

13. See. in convenient near-juxtaposition, E.J. Coates, "Notation in Classification" and J.E.L. Farradane, "Classification and Mechnical Selection. In International Federation for Documentation, Proceedings, International Study Conference on Classification for Information Retrieval. London, Aslib, 1957, pp. 51-64, 100-102, 65-69, 106-108.

14. For the most compendious statement by its originator, see Ranganathan, S.R. Colon Classification. (Rutgers Series on Systems for the Intellectual Organization of Information.) Vol. 4. New Brunswick, N. J., Rutgers Graduate School of Library Service, 1965, pp. 117-130.

15. As so often stated and clarified by Silvio Ceccato; cf., e.g., "Concepts for a New Systematics." In Proceedings, International Symposium on Relational Factors in Classification. (To be published by Pergamon Press, Oxford.)

16. The variety of subsumptions deduced in the present writer's paper "Categories and Relators: a New Schema," Revue Internationale de la Documentation, 32:136-144, Nov. 1965, shows this clearly enough; genus/species is only one of a whole family of such relations.

17. See, e.g., Haykin, David J. Subject Headings: a Practical Guide. Washington, U.S.G.P.O., 1951, pp. 1-11.

18. This conclusion is at least implicit in the present writer's discussion of the groups of subdivisions under the heading "Art"; see "Approaches to Library Filing by Computer." In Herbert Goldhor, ed., Proceedings of the 1966 Clinic on Library Applications of Data Processing, Urbana, Ill., University of Illinois Graduate School of Library Science, 1966, pp. 47-90.

19. Perreault, J.M. "Towards Explication of the Rules of Formation in UDC". In On The Perreault Schema of Relators and on the Rules of Formation in UDC (FID/CR Report no. 4). Copenhagen, Danish Centre for Documentation (in press).

20. As given in Universal Decimal Classification . . ..3d Abr. Eng. Ed. London, British Standards Institution, 1961, p. 10 (=B.S. 1000A = F.I.D. No. 289); the substantive codes given here are drawn from the same edition.

21. It will be seen that this is Fill's use of the square bracket, rather than its more common use as an "intercalator"; see Fill, Karl, Einführung in das Wesen der Dezimalklassifikation. Berlin, Beuth, 1960, pp. 20-21; see also footnote 19.

22. Metcalfe, John W. Subject Classifying and Indexing of Libraries and Literature. New York, Scarecrow Press, 1959, pp. 152-155.

23. Or, as the Dutch study-group "Grondslagen UDC" (affiliated with the Fédération Internationale de Documentation) says, in its report D66-196 (May 10, 1966), "(1) The Committee has diagnosed an increasing need of a more exhaustive consideration of unlocking potentialities through the UDC (depth classification or facet classification). (2) The Committee is of the opinion that the most effective way is to extend relations, rather than reconsidering or extending hierarchy."

24. Grolier, Eric de. A Study of General Categories Applicable to Classification and Coding in Documentation. Paris, UNESCO, 1962.

25. See e.g., Ranganathan, S.R. Classified Catalogue Code, with Additional Rules for Dictionary Catalogue Code. (5th ed. with A. Neelameghan). London, Asia, 1964, pp. 71-73.

26. An informative table of user-institutions is to be found in Mills, Jack, Guide to the Universal Decimal Classification (UDC). London, British Standards Institution, 1963, pp. 115-128. (=B.S. 1000C = F.I.D. No. 345).

27. See Freeman, R.R. Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System: an Introduction . . . .(NSF Grant GN-433). New York, American Institute of Physics, 1965 (Documentatin on Research Project Report No. AIP/DRP UDC-1).

28. See Freeman, R.R. Modern Approaches to the Management of a Classification. New York, American Institute of Physics, 1966 (Report No. AIP/UDC-3). Read to the 1966 Congress of the International Federation for Documentation. The amended goal will probably be something short of a full edition for all classes.

29. Caless, T.W., and Kirk, D.B. "Exhaustive Machine Searching with UDC." Paper read to the 1966 Congress of the International Federation for Documentation.

30. The question of how we know, if we are unsuccesful in our attempt to find precisely what we want, whether our further efforts are getting us any closer to it, will be explored more thoroughly in the present writer's forthcoming book, The Idea of Order: An Essay in Bibliographical Systematics.

VITA

Jean M. Perreault is a Lecturer in the School of Library and Information Services, University of Maryland, College Park, Maryland.

He was born in 1931 in Kansas City, Mo., received his B.S. there from Rockhurst College in 1952, an M.A. in philosophy from Marquette University (Milwaukee) in 1957, and the M.A. (L.S.) from the University of Wisconsin (Madison)in 1959. He was with the Milwaukee Public Library from 1956 to 1963, when he became first Chief of Cataloging and then Head, Information Retrieval Division, Florida Atlantic University, Boca Raton. In 1965 he began to teach at Maryland, and to work towards a Ph.D. in philosophy.

Mr. Perreault has published several papers on classification and its relation to information retrieval, and is now editing the proceedings of the International Symposium on Relational Factors in Classification (to be published this year by Pergamon, Oxford), of which he was General Chairman. He has served as consultant to private corporations as well as to such government agencies as the Public Health Service and the Environmental Science Services Administration.

* *

* *