

A study of general categories applicable to classification and coding in documentation

by Eric de Grolier

Professor at the Institut National des Techniques
de la Documentation (Conservatoire National
des Arts et Métiers, Paris);
Co-Chairman of the FID/CA Committee (General
classification theory) of the Fédération Internationale
de Documentation

Unesco

A study
of general categories applicable to classification
and coding in documentation

Published in 1962 by the United Nations
Educational, Scientific and Cultural Organization
Place de Fontenoy, Paris-7^e
Printed by The Ysel Press, Deventer, Holland

Translated from the French text (March 1960)
by the National Science Foundation, Washington;
it has been revised by Anthony Thompson, M.A., F.L.A.,
who also translated the author's additions (1961)



© Unesco 1962
Printed in the Netherlands
NS.61/IV.6/A

Contents

0	<i>Introduction</i>	9
01	Origin and character of the present study	9
02	Definitions and specification of terms	13
1	<i>General categories and relationships in general classification systems</i>	
10	Before the Universal Decimal Classification (UDC)	17
11	The Universal Decimal Classification	18
12	From the UDC to the Colon Classification system: the classification systems of the Library of Congress and of Bliss	42
13	The Colon Classification	44
14	After the Colon Classification	57
2	<i>General categories and the expression of relationships in specialized classification and coding systems</i>	61
21	France	61
211	Serge Tchakhotine	62
212	Z. Dobrowolski	63
213	The Desaubliaux code for the Carrel Foundation	64
214	Specialized classification systems inspired by Cor-donnier	65

Contents

215	The Robert Pagès coded analysis	69
216	Dr. Jacques Samain's coding system	75
217	The French Atomic Energy Commission's classification, and research on coding systems	75
218	Research by Jean-Claude Gardin	82
219	The Isambert classification system for the sociology of religions	89
22	Netherlands	89
23	Germany	91
24	Great Britain	92
241	J. E. Holmstrom	92
242	E. G. Brisch	93
243	J. E. L. Farradane	93
244/246	The Classification Research Group	94
244	D. J. Foskett	94
245	B. C. Vickery	96
246	Miss Kyle	97
247	J. Seymour	100
248	P. R. P. Claridge	100
249	W. S. Batten and G. M. Dyson	101
25	Canada	101
26	United States of America	104
260	Up to 1945-46	105
261	Calvin N. Mooers	106
262	The American Society for Metals and the work of J. W. Perry	107
263	Research at the US Patent Office	122
264	IBM research programme. H. P. Luhn	131
265	Mortimer Taube and the 'Uniters'	132
266	Miscellaneous specialized coding systems: in fields other than chemistry	133
267	Miscellaneous specialized coding systems: in the field of chemistry	134
268	Techniques aimed at using a more or less standardized form of natural language	134
269	Two special studies	138
27	Soviet Union	139
28	India	142
29	Japan	142
3	<i>General categories and the expression of relationships in natural languages, and experiments in international auxiliary languages</i>	143
31	Studies on categories and relationships in natural languages	143
311	Germany	144

Contents

312	France and Switzerland	145
313	Denmark	147
314	Great Britain	148
315	United States	148
316	Italy	149
317	Poland, Hungary, Czechoslovakia, Rumania, USSR, Yugoslavia, China	149
318	Some types of linguistic categories	151
319	Related problems	152
32	Categories and relationships in 'artificial' languages	155
4	<i>Problems of symbolization and notation</i>	159

Notes

Introduction	161
Chapter 1	166
Chapter 2	173
Chapter 3	205
Chapter 4	246

0 Introduction

01 Origin and character of the present study

The rapid development of information retrieval methods since the last war—together with the development of systems using mechanical or electromechanical means (punched cards) for such purposes followed later by electronic retrieval (computers)—have led to transformations in the structure of classification or coding systems for books and documents. Generally speaking, document specialists have tended to elaborate a large number of different codes, usually totally unrelated to each other, each conceived to meet the special documentation requirements of a particular organization. This is farther than ever—at least to all appearances—from the original intention in 1895 of the founders of the Institut International de Bibliographie (now known as the Fédération Internationale de Documentation), i.e., the standardization of classification methods.

It soon became evident, however, that this represented merely a transitory phase, preparing the way towards an ultimate stage during which it was hoped that this 'anarchy' would be eliminated, and replaced, if not by a single, new, universal-ly recognized standard, then at least by certain standardized

elements from the existing systems. These have revealed, in fact, a certain number of rather general practical procedures, resulting from common requirements and conditions which were more or less imposed upon them by characteristics peculiar to the new 'machines' (this term is used here in its broadest sense) for the retrieval of documents and the information contained therein. Such were, for example: the breakdown of complex subjects into simple factors (or, at least, less complex); the transition from highly hierarchical classification systems to less hierarchical systems;¹ 'the realization of the necessity of expressing not only the elementary terms used in the analysis of documents, but also the relations between them.

Gérard Cordonnier, in France, seems to have been among the first to emphasize the potential advantages of a synthesis of special codes.² In 1950, Mortimer Taube, at the conference on bibliographic organization held at the University of Chicago, presented a report on the 'Functional approach to bibliographic organization', in which he formulated, in his conclusion, the idea that Unesco could encourage the development and standardization of various 'categories' used in specialized codes, particularly those likely to apply to fields other than those for which they were originally intended.³ A similar point of view was expressed by L. I. Gutenmakher during discussions which took place in the USSR in 1952 at the Academy of Sciences (and which led to the creation of the Institute of Scientific Information).⁴ In England, following the Conference on Scientific Information, convened in 1948 by the Royal Society, the latter appointed a committee for the study of the problem of scientific classification; in February 1952, a group of librarians and information officers constituted itself under the name of Classification Research Group and submitted to Unesco a report emphasizing, in its conclusion, the importance of research into establishing a standard classification system.⁵

Recommendation No. 19 of the International Conference on the Analysis of Scientific Documents, convened by Unesco in 1949, had proposed that Unesco might sponsor the development of a standard classification system and assist in the establishment of a standardized international code for the mechanical retrieval of documents; at its first meeting, held in Paris in February 1954, the International Advisory Committee for Documentation and Terminology in the Pure and Applied Sciences (IACDocTerPAS) had concluded that neither the Universal Decimal Classification, nor any other already established classification, could be considered *a priori* as the most satisfactory method for the coding of scientific documents and recommended that a comparative study be undertaken of existing systems.⁶ A preliminary report was

prepared by us, in consultation with the secretariat of IACDocTerPAS, and was published on 1 June 1955. One of the conclusions reached stressed the importance of organizing research into the more general 'categories' (logical characteristics, characteristics of space, time, form and the various forms of 'action' in their broadest sense) which recur in all classification or coding systems applied to special fields of study or to special scientific and technical subjects.⁷

From 13 to 17 May 1957, an International Study Conference on Classification for Information Retrieval brought together a group of about forty experts from seven countries, which met at Dorking under the auspices of the Fédération Internationale de Documentation and through the good offices of Aslib, the Classification Research Group and the Library School of the University of London. This conference at Dorking may be considered as having marked an important turning point in the research on the standardization of coding systems; without regard to any preconceived ideas, it studied objectively the various conditions of such a standardization. It adopted a series of recommendations which included, notably, a description (point 9) of four research projects which it deemed particularly desirable, the first of which related to the 'preparation of classification tables for the more general categories applicable to a diversity of subjects (analytical, morphological, spatial, lists of properties, values, materials, etc.)'.⁸

During its meetings in Paris (16 and 17 September 1957) the FID/CA Committee resumed its study of the problem, and among the conclusions adopted (subsequently confirmed by the FID Council) appeared a suggestion on the desirability of seeking 'in co-operation with interested international organizations, such as Unesco', means of subsidizing a survey aiming at the 'establishment of a systematic coding of the most general concepts which recur in all fields of study: 'logical' concepts in the broadest sense, morphological, concepts of space, time, position, movement, etc.' This suggestion indicated that essential preliminary research would consist of reviewing and comparing the various attempts already made toward such a codification, e.g. those made by Messrs. Kervégant, Pagès, Perry, Vickery, the US Patent Office, etc., and to ascertain from a study of the various general classification schemes (particularly the UDC) and the special systems, how they have attempted to codify these concepts. The resolution added that it would also be necessary to study scientific and technical terminology from this point of view, and mentioned in this connexion the work begun earlier by L. Couturat.

Finally, at its third meeting in Paris (23-25 September 1957) the IACDocTerPAS Committee, upon being informed of the

resolutions adopted by the FID/CA Committee, and of the report which we had prepared for the latter's meeting, mentioned above,⁹ recommended that 'a contract be prepared (by Unesco) for a survey of sufficiently broad categories to be used in classification and coding systems in numerous different fields of knowledge—logical and morphological concepts, space and time, position and movement—and for the preparation of a report on the subject, to be presented at the International Conference on Scientific Information in Washington, in November 1958, or to be published by Unesco'.

This report is presented here.

We apologize, at the outset, for its somewhat incomplete and provisional character—although it has taken almost two years to write. These shortcomings are due in part, of course, to our own personal deficiencies in the face of such a vast subject, which would require, for adequate treatment, to be handled by a mind of infinitely greater breadth and range than ours, but it should certainly also be admitted that our task has been rendered particularly difficult by the rapid evolution which has taken place in this field in the last few years. A great deal of research has been undertaken in various countries, dealing more or less with the questions we are studying; it is not always an easy matter to keep abreast of such developments adequately, nor to obtain documents which relate to them.

Following Cordonnier, however, we had recognized at the outset the relationships which existed between the establishment of a general classification, the creation of a 'common language' for the machines designed for the retrieval of information, and that of an international auxiliary language; we had already, for a number of years, tried to follow the development of research in the latter field. This research, however, has been the subject of a complete revision in view of the spectacular strides made in the study of mechanical translation.

As early as 1951, Yehoshua Bar-Hillel—one of the pioneers of mechanical translation—introduced the concept of a 'universal, or at least general, grammar, perhaps even the construction of a whole artificial exchange language' and of a language standardized for the writing of scientific and technical analyses in such a manner that the 'sentence-pattern translation might easily and quickly be applied, perhaps not directly into any other language, but first into some exchange language, natural or artificial.'¹⁰

Similar ideas were expressed by several researchers: James W. Perry,¹¹ Erwin Reifler¹² and Luitgard N. Wundheiler.¹³ And, although the first meeting held concerning mechanical translation, at the Massachusetts Institute of Technology in the spring of 1952, agreed as to the desirability of establishing a priority

for experiments in the mechanical translation of a natural language directly into another natural language, it had by no means excluded the possibility of an 'inter-language'.¹⁴

It seems that, in the Western world, developments then tended towards other ideas, and that much greater scepticism now prevails regarding any attempt to use a 'meta-language' as an intermediary for automatic translation. Booth, Brandwood and Cleave, in the very brief chapter of their work of 1958 devoted to multilingual translation, considered general agreement to be impossible concerning the creation of a universal language, questioned its usefulness for mechanical translation, and ended by advocating the 'acceptance of normal or slightly modified English as the language to be selected as a basis of operations'—recognizing, however, that such a suggestion would naturally be most unpopular, both with 'cranks who endeavor to advocate the adoption of their own particular artificial language' and with Russians and Germans.¹⁵

Nevertheless, the conference held in Cleveland (6 to 12 September 1959) by Western Reserve University was named 'An International Conference for Standards on a Common Language for Machine Searching and Translation', and, on the 'Western' side, G. Cordonnier presented a report on a 'Meta-language for human communication, translation, and machine searching', while A. F. Parker-Rhodes spoke on 'Some recent work on thesauri and interlingual methods in machine translation';¹⁶ the latter relates to research by what can be referred to as the 'School of Cambridge' (Cambridge Language Research Unit) which stresses a common method (use of a 'thesaurus'-type dictionary) both for information retrieval and for mechanical translation purposes.¹⁷

At the same conference, a Russian report was presented, indicating that, in the USSR, among the various research groups in the field of mechanical translation, some of these are engaged in a search for an 'ideal pivot-language'.¹⁸

It can be anticipated, therefore, that in the near future a close liaison will be maintained between the work in progress on improving mechanical translation and the efforts towards the mechanization of information retrieval.¹⁹ In view of this, our present study should certainly have examined in greater detail and greater depth the purely linguistic aspect of the problems studied; but this will have to be the subject of a later study.

02 Definitions and specification of terms

Among the research programmes indicated as necessary in the conclusions and recommendations of the Dorking Conference,²⁰

was 'the establishment of a unified and systematized terminology in the field of the theory of classification itself'. Unfortunately, this unified terminology does not as yet exist—one might even say exists less than ever.

This is not the place to work out such a terminology in detail;²¹ we shall content ourselves with definitions of the meaning which will be subsequently attached to certain terms used in the present study.

First, it is necessary to define the use of the terms 'coding' and 'classification'.²²

The word 'code', indicates the *Nouveau Larousse Universel*—apart from its accepted meaning of 'a collection of laws' (arranged in order) or 'that which serves as a rule'—applies to a 'dictionary of conventional signs for telegraphic transmission in an agreed or abbreviated language'. More broadly Guilbaud defines a code as a 'system of conventional signs which permit faithful translation between unevenly developed groups of signs or symbols',²³ such 'groups' designating here a body of signs or symbols used in the transmission of information.

'A language', writes Joshua Whatmough, 'is a code in which messages are transmitted':²⁴ it should be defined that language, in its most usual sense, is a communication medium based 'on the association of thought content with sounds produced by speech'.²⁵ Whatmough adds that the 'systematic symbolism' of the language 'can be transformed into other systems', electric (as in the telegraph or the telephone) or electronic.

Benoît Mandelbrot elsewhere describes 'the very broad classification systems and certain segments of the current vocabulary, which are analogous' as 'inferior forms of language':²⁶ the classification systems which are the object of his study are 'natural systems' of the Linnean type, reducible to taxonomic dichotomic trees and follow the law of Willis.

Following this trend of thought, one might consider that the term 'coding' is a very general term, to be used for any type of designation of elements intimated (ideas, concepts) by means of a finite series of predetermined special signs; in language (natural) this series of signs being composed of phonemes—which may be translated more or less satisfactorily by graphic signs. 'Classifications' would be a special type of coding, the freedom of the coder being restricted by special conventions, and above all by the obligation imposed upon him to follow a certain hierarchical order (whether this order is 'strong' or 'weak').

The term 'code' has been considered, however, in its narrowest sense in the field of the use of machines for information retrieval. It is thus that H. P. Luhn, in distinguishing three categories of codes (letter codes, word codes, cryptographs), is con-

cerned primarily with translating the terms of current language into a series of signs which can be handled by a machine.²⁷

Proceeding further in the same direction, and certainly too far towards a specialized sense of the word, Martin Scheele writes that a code is 'a determinant for the location of holes on a punched card'.²⁸

Finally, and subject to a more detailed analysis, it would seem that the definition of 'code' may be accepted as given in the 'Glossary and subject index' mentioned in note 21: a system of symbols for the representation of information and rules governing their combination;²⁹ this is how we shall interpret it here. Furthermore, we shall interpret 'classification' as a coding system within which the 'words' of the code (series of symbols indicating a concept, or semantemes) are subject to certain order relationships.

It is now necessary to try to define the word 'category'. Here unfortunate confusion prevails in the literature of our subject. At times the term is taken in its broadest and most general sense: for instance, Wildhack considers it as a synonym of 'point of view' according to which a subject can be divided;³⁰ D. J. Foskett considers it as a synonym of the term 'faceted' brought into fashion by Ranganathan, and writes that 'facet analysis' consists in an analysis of a subject in its entirety 'into a certain number of facets or categories of things; within each category, the subject headings enumerated all possess the same relationship vis-à-vis the subject in its entirety'.³¹ At other times, 'categories' is understood in the sense of 'general concepts' (this is the third sense (C) of the word in the *Vocabulaire de la philosophie* of Lalande). The 'Glossary and subject index', already mentioned, defines the word category as follows: 'A concept of high generality and wide application which can be used to group other concepts'.³² From his point of view, B. C. Vickery, inspired by L. Wood's definition, understands 'conceptual categories' to mean 'concepts of a high degree of generality with a wide area of application elaborated by the mind in referring directly or indirectly to empirical knowledge and utilized by the mind in interpreting such knowledge'.³³ In Ranganathan's terminology, here again extremely specialized, the term 'fundamental category' is used in a particular sense, 'each facet of any subject, as well as each division of a facet, is considered as a manifestation of one of the five fundamental categories'.³⁴ Although this idea can doubtless be traced to philosophical or even mystical traditions, in Ranganathan it seems related to a very practical preoccupation, that of 'insuring a uniform sequence' of the 'facets' under the various subjects.³⁵

In the present report, we shall not limit ourselves to an examination of such 'fundamental categories', nor, indeed, to any

other basic systems of categories for the grouping of code-words in formal classes after the model of grammatical categories, such as that used by Leroy and Braffort.³⁶ We shall try, in a more general manner, to examine all the procedures which, by means of a coding scheme used for the retrieval of information contained in documents, would obviate the weightiness and growing complexity of simple *enumerations* of terms, in order to indicate the *relationships* between such terms. We do not attempt to conceal the rather vague as well as doubtless unsatisfactory character of this formula, which we hope, nevertheless, will be clarified somewhat by a concrete discussion of the special methods described in the following chapters.³⁷

In the following chapters, first those procedures will be examined which are used for the expression of general categories and general types of relationships and which are to be found—in a more or less embryonic or confused state—in a certain number of general library classifications, and also in special coding systems. A third portion of our report will relate, for comparative purposes, to similar procedures used in 'natural' or 'artificial' languages; we beg our readers to consider it simply as a preliminary rough outline. The problems of symbols and notation raised by the expression of general categories and 'relationships' will be briefly dealt with at the end of this report.

We should have liked to conclude this report with a sample of systematization, for discussion, of at least a certain number of general categories. But although we have drafted, for a number of years, various attempts at 'categorization' of this particular type,³⁸ we judged none of them to be sufficiently satisfactory to warrant even preliminary publication. Accordingly, the present report lacks a positive conclusion. We fear it will have raised more problems than it has been able to solve, and also often left discussion in mid-air; our sole ambition is that it may serve as a frame of reference for subsequent studies, or, more modestly, that it may facilitate perhaps for other researchers, better equipped and better prepared than we were, more fruitful explorations.

The present work was issued provisionally as a duplicated document, in September 1959 for parts 1 and 2, and in March 1960 for parts 3 and 4; in March 1961 it was withdrawn, and we have since added a number of notes and bibliographical references; the original text has also been slightly retouched, as a result of various observations received, and for which we wish here to thank the authors (especially Dr. Burton W. Atkinson).

1 General categories and relationships in general classification systems

10 Before the Universal Decimal Classification (UDC)

Although it is not necessary to relate the history of the subject under study, it is of interest to note that, as early as the seventeenth century, the classification schemes of a few large libraries had begun to use in their schedules what we have now termed 'parallel' divisions.¹ Instead of each subject being divided in its own particular manner, a common principle was adopted and the same order was fairly closely followed in the subdivision of different classes. These related generally to geographical or historical divisions.²

Melvil Dewey used this parallel method rather widely in the Decimal Classification, first published in 1876, especially for the divisions of literature, philology, and history (classes 800, 400 and 900). He seems to have been the first, however, to clear a new stage and to introduce the principle of divisions 'by transfer' (a term used by us in the study mentioned in Note 1), primarily for what he called 'generalities'—i.e. for subdivisions based on the characteristics of the document, either according to the point of view of the author (bibliographical, theoretical, or historical viewpoint) or according to the form of the document (monograph,

dictionary, periodical). This was, however, a slightly later development and dates actually from the second edition of the Decimal Classification (DC) (1885); it was followed in subsequent editions, where more and more frequently one encounters the 'divide as in...', which indicates these divisions by transfer. Dewey may have been influenced, however, by his colleague and rival Charles A. Cutter, whose Expansive Classification had been the object of a very brief review in 1879³ and had been extended to its 'fifth expansion' in 1882, the first relatively complete edition being dated 1891-93.⁴ We are undoubtedly indebted to Cutter for the first so-called 'common subdivisions', separated from the table of general divisions, which he used to indicate literary form and to which is added what he referred to as the 'local list', i.e., a list of geographical divisions, applying to any subject.

We shall dwell very briefly here on James Duff Brown's 'subject classification', although it was published in 1906, namely a year after the completion of the first edition of the Universal Decimal Classification by Otlet and La Fontaine (1899-1905), because its origin dates back to 1898.⁵ It includes divisions of form, place, languages and chronological divisions, and its 'categorical tables' list an extensive series of 'common stand-points', by which any subject can be subdivided. As observed by Phillips, quoted by Vickery,⁶ Brown thereby avoids most of the 'multiple locations' which are a characteristic trait of the Dewey classification; the 'tables of categories' indeed allow Brown to include, once only in general, a specified subject in the main classification plan, and to pinpoint the various 'aspects' under which the subject can be considered by the addition of an affix, an auxiliary term taken from the table of categories.

11 The Universal Decimal Classification

It is very possible that Brown was influenced by the work of Paul Otlet and Henri La Fontaine who, about 1893-94, had begun to develop the Dewey Decimal Classification for the detailed classification not only of books but also of articles in periodicals.

Applied to such use—for which it had not been conceived—the DC showed its lack of flexibility. Otlet and La Fontaine (their respective parts are difficult to ascertain, although it seems that the basic inspiration was provided by Otlet), basing themselves on the Dewey principle of 'divide as', conceived the idea of distinctly separating the common subdivisions in the notation of the classification, 'introducing' them by special symbols, distinct from the digits of the main decimal notation; these symbols were derived from the arsenal of punctuation symbols, to which were added a small number of mathematical symbols. There again, Dewey had

been the forerunner inasmuch as he used a period (full stop) to separate the number into two parts, between the first series of three digits corresponding to the first thousand divisions and subsequent subdivisions; although the period's usefulness was merely that of facilitating the reading of the classification numbers, however, when followed by a zero, it denoted subdivisions of form (bibliographies, treatises, dictionaries) or geographical and historical forms (.09).

The UDC therefore contained 'tables of common subdivisions' (this term is perhaps derived from Cutter) for places (indicated by parentheses), chronological divisions (between quotation marks), for languages (preceded by the symbol =); for races (between parentheses, the first of which followed by the symbol =); for bibliographical form (between parentheses, the first digit after the first parenthesis being a zero). To these were added the so-called divisions of 'point of view', indicated by a period followed by a zero, or 'analytical divisions', preceded by a hyphen.

It is perhaps more important to note that the UDC was the first document classification system to introduce a relationship sign—the colon (:). This represents a 'general relationship' as indicated by Otlet himself? non-differentiated, meaning merely that the subject is studied in its relationship to another. Numbers linked by the symbol : are, moreover, reversible; 02:334 may be indicated for the subject 'co-operation in the library sphere' (02 meaning library, derived from class 0 generalities, and 334 meaning co-operation, derived from class 3 social sciences, and a subclass 33 economy), or else 334:02.

The use of the sign 'general relationship' was never, however, intended by those responsible for the evolution of the UDC (Otlet and La Fontaine first, followed by F. Donker Duyvis) to replace the normal subdivision method, which remained—as in Dewey—a direct division of a number corresponding to a concept of a certain level of generality into numbers corresponding to more specialized concepts by the simple addition of a supplementary decimal figure, on the assumption that the 'more specialized' concept was related to the 'more general' concept by a hierarchical relationship. Consequently, the 'general relationship' was added to the only type of relationship hitherto known in library classification, i.e., the hierarchical relationship between genus and species (of inclusion), merely as an auxiliary, with a view to providing greater convenience, more flexibility in certain cases, and definitely not to serve as a basis for the classification—the latter retaining essentially its traditional hierarchical character. In fact, in many cases where a division *could* have been indicated by : a direct division was preferred.

In short, the UDC provides the means of expressing two types of relationships between subjects: the hierarchical relationship, or inclusion (similar to all the other classification schemes which preceded it—and, in fact, a large number of the schemes that followed still know of no other), and a 'general' relationship, non-differentiated, which covers all other types.

D. Kervégant⁸ considers that the addition symbol (+), used in the UDC to indicate that a document relates to two subjects, separate and distinct from each other in the tables of classification (for instance a study on sculpture 73, and painting 75, which would be shown as 73 + 75) must also be included in the methods expressing relationships. It corresponds to the logical operation of 'reunion', but its use in the UDC is extremely limited.

The present author has attempted to clarify the relationships indicated by the sign : by adding to the latter diversely oriented arrows followed by an order number, as shown in the following table.

1. *Appurtenance (belonging)*

- 11 → Inclusion, implication
- 12 → Parts, organs
- 13 → Components, constituents
- 14 → Properties, attributes
- 141→ " " physical
- 142→ " " chemical
- 143→ " " biological
- 15 → Aptitudes, predispositions

2 *Process*

- 21 → Action: acting on (subject), affected by (object)
- 211 → Favourable (stimulation, increase)
- 212 → Unfavourable
- 2121→ Delay
- 2122→ Inhibition
- 2123→ Destruction
- 21 ← Interaction
- 211 ← Favourable (symbiosis)
- 212 ← Unfavourable (antagonism, competition)
- 22 → Operation, means used: process (subject), product, result (object)

3. *Dependence*

- 3 → Causality, origin, etc.
- 31→ Causality; cause (subject), effect (object)
- 32→ Origin: originating (subject), arising from (object)
- 33→ Condition, requirements: conditioning (subject), conditioned (object)
- 3 ← Interdependence
- 31← Correlation
- 32← Association
- 33← Combination, synthesis

4. *Orientation*

- 41→ Aspect, particular case
- 42→ Application
- 43→ Use

5. *Comparison*

- 51 → Resemblance, likeness, similarity
- 511→ Analogy
- 512→ Equality, identity
- 52 → Dissimilarity, unlikeness
- 521→ Difference
- 522→ Opposition (of character)

The negation of a relation may be represented either by putting a zero before the number used, or by putting over it the sign used for this purpose in logistics, the dash.

This interesting proposition—and we shall encounter later similar views (by Perry and Kent; Gardin; Leroy and Braffort)—was not accepted by the Central Committee on Classification (FID/CCC) which watches over the development of the UDC. Apart from the resulting complication in the use of such classification numbers in printing or typing, the principal reason was probably that the adoption of such a method of detailed classification of relationships would completely upset the established UDC order. In fact, a large number of existing classification numbers include, for instance, appurtenance (property) relationships or ‘orientation’ (application) relationships, obviously confused—through the lack of special symbols—in the normal hierarchical relationship, but which are nevertheless present. It may be considered, for example, that the considerable detail under class 6 applied sciences is the equivalent of 5 :→42→ and of its divisions in a decimal ‘classifying language’ which would apply Kervégant’s application relationship. ‘Properties’ are expressed, in class 66 chemical industries by the division 66.016 and its subdivisions, for non-chemical properties, and by 66.014 for chemical properties, and other ‘properties’ are found in various divisions.⁹

If we return to the UDC as it stands at present and without regard to common subdivisions of form, language, or race, which relate either to relative characteristics, not of the *content* of documents, but to certain external aspects of their subject (form of presentation, language used), or which relate to divisions applying only to certain classes (forming a part of anthropology) and not to the classification scheme as a whole, it would seem desirable first to examine in some detail the UDC’s method for treating the general notions of space and time.

An important preliminary observation should be made here: the UDC is not a classification scheme by ‘objects of study’,

but a classification by branches of knowledge or human activities. This results in any 'object of study' being classified in all those places where it may be studied by any particular science or technique; this dispersion, naturally, is extended still further for such general 'objects of study' as *space*.

The abstract scientific study of *space*, geometry, appears in the UDC as a subclass assigned to mathematics 51, forming part of main class 5 entitled 'mathematics and natural sciences'; it is at 513. The philosophical study of space is found at 114, a division of 11 metaphysics, general cosmology. One finds under 531 (applicable up to 534 — divisions 531 to 534, being part of physics, cover mechanics) some 'analytical' subdivisions identified by a hyphen:

- 1 lines
- 2 surfaces
- 3 space
- 4 multidimensional space
- 9 non-Euclidian space

Under class 574/578 general biology (the symbol / represents a grouping of classes in which the classification numbers run consecutively), other analytical subdivisions by hyphen are used: -181 for sizes, dimensions (-181.12 length and width; 181-13 height) and -182 for 'biogeometry' (comprising a single division only, -182.2 symmetry and asymmetry).

Under 621, a division of 62 engineering, technology and industry in general, also included in 6 applied sciences, medicine, technology, will be found a division 621 entitled, in a recent abridged trilingual edition FID 277 (1958), 'mechanical and electrical engineering', but F. Donker Duyvis' memorandum F58-75 (4 September 1958) states that this denomination is in error and should be replaced by 'mechanics and applied physics in general'; there will be found here an impressive series of analytical subdivisions by hyphen, which apply furthermore to all the 62/69 divisions, in so far as they relate to the characteristics of machines: -181 relates to the characteristics of machines and equipment according to their size (-181.12 length and width, -181.13 height, -181.2 large machines, -181.4 small machines and pocket machines). Again in these subdivisions by hyphen under 621, -9 is entitled 'motorized machines in general', but its divisions, limited to two, relate to topological notions: -962 parallel coupling, -964 series coupling. It contains at least another example of utilization of -181 for dimensions, under 575 heredity, genetics. On the other hand, in anthropology they are specified by direct divisions under 572.512 body dimensions and proportions in general; under 572.543/544

dimensions of the head and the face; under 624.032 dimensions of buildings; and 625.73 highway shape and dimensions. The dimensions of transport vehicles other than those on rails are found in 629.1.071.

The following other divisions relating to space in the principal UDC tables or their analytical subdivisions by hyphen¹⁰ may be mentioned: 301.16 social environment, which is a part of 301 general sociology, and the 'ecology' classes in botany 581.5, zoology 591.5, and agriculture 631.95.

Under the common subdivisions themselves, space does not appear in its general aspect. *Place* only is given, which Lalande's *Philosophical Vocabulary* defines as 'area occupied by a body, distinct from surrounding space and considered as a part of space', which is therefore a more restricted notion. 'Common subdivisions of place' are symbolized by numbers between parentheses, the first figure in the latter extending from 1 to 9 (should the first figure be a zero, then the number would relate to an entirely different category, that of so-called 'form' divisions, which relate only to the form of documents, considered either under their intellectual aspect—form of presentation—or according to their material aspect).

The more general notions of place are found in the analytical divisions as (- . . .), which relate to location divisions (1) to (9) 'for further detailing, limits, boundaries, political groupings, etc.' states the abridged trilingual edition in its French section (p. 35); in fact, a certain number of these are more general in character and could be applied to extraterrestrial locations—at least theoretically: as in the case of divisions (-19).

- (-191) relative position. Interior, internal
- (-191.2) central, middle
- (-191.6) eccentric
- (-192) peripheral, encircling
- (-194) exterior, external, outside
- (-194.2) adjoining, nearby, neighbouring
- (-194.4) half-way or equi-distant
- (-194.6) distant
- (-195.2) before, in front
- (-195.6) behind, in the rear
- (-196.2) to the right, right side
- (-196.6) to the left, left side
- (-197.2) down, toward the bottom, lower side
- (-197.6) up, toward the top, upper side
- (-198) other relative locations
- (-199) without specification of place

It must be added that all the above subdivisions are recent (memorandum PE 492, that is 1951 or 1952), and that the 1958

abridged trilingual edition did not consider it necessary to indicate all of them: it ceased at (-194.2). Their application is therefore restricted.

A proposal by Dr. Fill (note P.727 of 26 January 1961) suggests modifying division (-06) and using (14) and (141) to (148) for the divisions of 'Orte nach geographischer Länge und Breite'—of places according to their latitude and longitude; thus analytical subdivisions would no longer be used for this purpose.

The other analytical divisions of place are intended to specify the political or administrative districts of a country (-2/-5): local, regional, etc., or to express certain concepts such as 'empire' (commonwealth) or 'colonies'; the (-6) division covers 'groups of states in wartime' (it seems that to date the UDC has not made any provision for groups of states in peace time . . .); (-77) indicates under-developed countries, but it does not seem to have any counterpart for developed countries. Curiously, there appears as a (-7) division 'spheres or various areas of activity of private enterprise' (according to the French text of the abridged trilingual edition, corresponding very nearly also to the German text, but the English text indicates 'for private and (semi-) public enterprise'), there seems to be nothing in the area of public activity, at least at that level; the (public 'domain' appears under 351.71 in the principal tables, a division which is a part of 351 activities of public services (under 35 public administration, administrative law, military arts and sciences), and includes fiscal law also.

At the end of the (- . . .) are again found three analytical divisions of 'place according to situation or origin' (this translation of the French text of the abridged trilingual edition is not too clear, but it was translated from a German text itself rather obscure: 'Der Ort als Schauplatz und Ursprung'; the English text is 'source and destination'): (-82) origin or source, (-85) destination, (-87) foreign, alien.

The general relation symbol (the colon) may be used for common location divisions, but in the restricted sense of 'place in its relations with other places', (43:/44) meaning the relations between Germany and France, without apparently being able to indicate the trend of such relations; a partial adoption of Kervégant's ideas might have been indicated here. It appears impossible, therefore, to express by the UDC such ideas as 'a Roman Catholic country' or 'a socialistic country' by combining the idea of location with a notion relating to an activity; on this particular point, as we shall see, the Colon Classification is more flexible.

The situation in so far as (2) is concerned—at least as far as we can see—is rather confused. In the abridged trilingual edition, the French text indicates 'physical places and environment'

but the English text gives 'physiographic designation', which tends to restrict the meaning. The 1934 German edition (third complete international edition, the only one at present available for the schedules in their entirety) gave the following general table:

- (203) in the air
- (204) in the water
- (205.5) in the light
- (206.1) in solids
- (206.2) in fluids
- (21) continents
- (22) islands
- (23) mountains (divided by countries; with subdivisions by .0 for low (23.01), medium (23.02), and high mountains (23.03))
- (24) interior of the earth
- (25) plains, deserts
- (26) oceans, seas
- (27) ocean currents
- (28) fresh waters

The abridged trilingual edition does not contain the (20) divisions; as to the remainder, it conforms to the above table. We have received, however, from Mr. Donker Duyvis a typewritten text dated 15 April 1958, 'bijlage brief Prof. Dr. Groen' [enclosure, letter from Professor Dr. Groen], which does not seem very compatible with this. Its main divisions are, in fact, as follows

- (202) cosmic space
- (203) atmospheric layers (with subdivisions, including (203.21) biosphere)
- (204) hydrosphere, under water (subdivided (204.1) submerged state and (204.2) afloat, plus two further divisions for the plankton layer and the bathysphere)
- (210.4) lithosphere in general
- (210.5) superficial layers in general
- (210.6) speleological strata in general
- (210.7) stratigraphic strata to be subdivided as 551.7 [in a letter dated April 1958, F. Donker Duyvis indicated (119) for stratigraphy]
- (210.8) layers between lithosphere and barysphere
- (210.9) barysphere, core of the earth

However, this is undoubtedly only a proposal.

It must be moreover noted that, in the main schedules are found—under division 551 'general geology, meteorology, climatology, stratigraphy, paleogeography'—divisions which practically duplicate most of those of (2), for instance: 551.14 crust of the earth, lithosphere; 551.16 core of the globe, barysphere; 551.4 surface of the globe (with divisions from which have been patterned

those from (21) to (28); 551.5 meteorology; 551.7 stratigraphy.

A similar duality is found with regard to 'celestial place': they form the object, on one hand, of common subdivisions under (*1/9), elaborated by Bigourdan for the 1927 French international edition,¹¹ and, on the other hand, of divisions within the main tables, in the section on 52 astronomy:

- 523.1 universe, world system (where is found the 523.14 sub-division for interstellar space)
- 523.2 solar system
- 523.3 moon
- 523.4 planets
- 523.5 meteors and zodiacal light
- 523.6 comets
- 523.8 stars
- 525 earth (in 525.7 is again found terrestrial atmosphere)

which are not particularly striking as far as very strict logical order is concerned.

Division (3) relates to 'places in the ancient world', whereas (4/9) relates to 'places in the modern world', the whole constituting (3/9) 'political locality'. This dichotomy, inherited from Dewey, is not without some disadvantages; already artificial in character in separations such as (37) Rome, ancient Italy, and (45) Italy (modern), it fringes on absurdity when one finds 'China, Cinarum-regio, ancient Japan' under (31), China ('modern'—but starting when?) under (51), and Japan (modern) under (520). We shall not dwell further on these divisions, certain distortions of which we have criticized elsewhere.¹² Within the 'common subdivisions of point of view' by .00—which, born of the necessity of applying the UDC to the classification of public archives, are, in fact, more specifically 'administrative' points of view, in a rather broad sense—is found a .006 division, the French title of which, in the abridged trilingual edition, is 'point of view of premises and establishments'; the German titles (Räume und Ortlichkeiten) and English (Space, site, accommodation) are more general in character. This division is inserted between .005 point of view of furniture, fittings and equipment, and .007 point of view of personnel.

The UDC's divisions concerning *form* should now be examined. In as much as this most equivocal word corresponds in fact to a well-defined concept, it is a spatial concept. Langevin defined the form of an object as 'the aggregate of simultaneous positions of all its points';¹³ the definition contained in Lalande's vocabulary read 'geometrical figure constituted by the contours of an object', and opposed form to matter, stating that this opposition could extend 'by metaphor to all analogous oppositions'. To the notion of form is closely linked that of *structure*,

which Lalande's vocabulary describes as 'arrangement of the parts which form the whole' (as opposed to their functions). The 'Gestalttheorie', defined by Lalande as applying to phenomena no longer merely consisting of sums of elements, but together constituting autonomous elements, possessing an internal solidarity, their own laws, and the behaviour of each element depending upon the structure of the whole, in fact concerns structures.

A uniform treatment of the morphological divisions is not found in the UDC and it would be a difficult task to furnish a complete list of all classes where such aspects may be found; a list of the principal examples should suffice. The morphology of language is found under 415.5. The shape of the earth is under 525.14 (in astronomy), however, one may also find under 551.11 (551.1 being general geology) a heading reading 'generalities on the shape of the earth'; 539.2 relates to the structure of molecular systems (together with their properties). In chemistry, chemical structure in general is found under 541.6; in colloidal chemistry 541.18, the .02 analytical division relates to structure, with, moreover, various subdivisions according to sections. An analytical subdivision by -162.2 applying to the whole of class 54, relates to crystalline form from the standpoint of crystallographic structure, to be divided as 548.1.02 (crystalline systems). In organic chemistry 547, analytical subdivisions by -0 apply to structural characteristics of organic liaison; these subdivisions are further developed for the different groups, for example for 547.53 benzene and homologues. 551.3 external geodynamics, contains the various types of facies; geomorphology being at 551.4, and structure of rocks at 552.122; however, in the 1944 supplement of the German edition, are found divisions 552.125.4 'Gestalt der Mineralien' [Form of minerals]. Under 572 anthropology, the configurations of the human body are distributed between 572.5 'somatology according to Professor Martin', and 572.7 morphology, 'merology according to Professor Martin'. 'General' morphology (i.e. living beings) is found under 576.2, together with general physiology, but morphology of cells is at 576.31, that of bacteria and parasites at 576.8.094, that of plants at 591.4 (the 4 may be transformed into -4 to serve the purpose of an analytical division of the various classes of systematic botany), and that of animals at 591.4 (with the same comment). Human anatomy is located under 6 medical, science, at 611 (separated, it will be seen, from 'somatology' and 'merology in the sense of Professor Martin', by the whole of biology); in 343.93, a direct subdivision of 343.9 criminology (itself a division of penal law), is the 'anatomy and physiognomy of delinquents'.

Under 621-4, will also be discovered an important series of analytical divisions with a rather general application, such as

those we encountered above (p. 22), concerning any and all categories of 'machines'. They bear the title 'external shape of semi-finished products' in the French section of the abridged bilingual edition¹⁴ and are derived from division 621.9 tools and machine tools, machining, cutting, etc.¹⁵ They should not, however, be used under 621.4 heat engines, except steam engines (under 621.43 internal combustion engines, the -4 relates to combustion and injection). To the latter restriction, already mentioned in the German edition of 1938-44, the abridged trilingual edition of 1958 has added classes 625.2/6 railway rolling equipment and 629.1 means of transportation other than railways; under this last heading, in fact, as early as 1938, -4 had been used in a different sense—that of use and destination of vehicles.

Deformations, in general, appear under 539.3 mechanics of elastic solids; those of apparatus are, in 621, under -75. But the calculation of deformations, in civil engineering (624), is found under 624.044 graphic and analytic statics. Deformations of the soil are located under 624.131.54 (soil statics and dynamics being at 624.131.5).

In various places is found the *action of giving form or shape* and the machines and methods which permit such action: under 621.7 metal work (forming); 621.97 machines for stamping, forging, compressing, moulding, presses; 678.027 shaping of rubber and plastics in general (however, preforming is found under 678.034, and shaping machines under 678.057); 666.1.032/037 for glassblowing, but spinning of glass is under 666.1.189.2.

Under 624.036 is found a division for 'buildings from the standpoint of their direction or their plan' (here again, the German text 'nach Richtung oder Grundform' [according to direction or basic form] or the English text 'layout and basic form', appear more desirable).¹⁶ 624.07 relates to 'elements of construction in general'; the English text 'structural elements' indicates more clearly that this relates to questions of structure.

There are divisions for the form or shape of ships (629.12.011.1) and those of aeroplanes (629.12.011.1), but division 011.1 for terrestrial vehicles is devoted to other aspects (generalities on chassis and framework) and we have been unable to discover another number available for the expression of the above notion.

Soil profiles appear under 631.47.

In 66 industrial chemistry there are divisions by -965 for the form of the raw materials.

Under 669, metallurgy, it is indicated that divisions 621-4 relating to form or shape can be adapted; the same applies to 674, lumber industry, and under 678 to the macromolecular substances industry. The note on p. 682 gives the divisions by -4 for the form of textiles.

In 691 construction, building, the 'exterior form or shape of elements of construction' is also indicated by the division 691-4, subdivided as 621-4; it is to be feared, however, that the application of these subdivisions conceived for a totally different purpose, will prove to be somewhat difficult.

Curiously enough there seems to be under 7 fine arts, only one analytical division by .08 for forms of musical works—but nothing for the forms of other arts. Under 72 architecture (trilingual abridged edition), in 72.01 aesthetics, theory, is found .013 proportions, harmony of measurements (which was .012.3 in the German edition of 1948). 'Literary forms', on the other hand, occupy the analytical divisions by hyphen of 8 in their entirety—but at this point we reach the limit of rational expansion of the meaning of 'form'.

All of the above, from a general standpoint, seems to lack coherence, at least in the eyes of an outside observer. The treatment of morphological and structural notions in the UDC seems to have been governed principally by tradition and the particular initiative of the various specialists who have developed such and such a part. Otlet, when using the (0 . . .) for the form of *documents*—and not for notions of form in the *content* of the latter—had doubtless committed an error and followed too confidently the way opened by Dewey. It will be noted, however, that the -4, first introduced for very particular purposes, tends to 'invade' other fields. Here, without doubt, exists a means for the UDC to 'standardize' the classification from the 'formal' standpoint.¹⁷

Let us now proceed to the treatment of notions of *time* in the UDC.

The notion of time in general appears in the main tables in two places: at philosophy (metaphysics) in 'cosmology' 113/129, under 115 time, duration, notion of space-time continuity (French text of the abridged trilingual edition; the German reads 'Verbindung von Zeit und Bewegung' [connexion of time and motion] and at physics under 531.11 bases of kinematics, where 531.111 is entitled 'dimensions, space, and time'. Implicitly, the notion is present in 530.12 principle of relativity. Time measurement is found in 52 astronomy, under 529 chronology, calendars, time measurement.

Chronological divisions in the UDC are usually obtained, as is well known, in a very simple manner, and even, it may be said, mechanically, by inserting, between quotation symbols, the extreme dates of the period under consideration, according to the Gregorian calendar. If the periods in question are not precise as to the number of years, an abbreviation is made to two or three

figures; if the periods are to be specified more accurately than by the indication of the year or years, then we insert, following the four figures of the year, two figures for the month, and two for the day (of the month). Conventional signs: “—” for ‘antiquity’, the pre-Christian era, “+” for the Christian era, “04/14” for the medieval period, “15/18” for the modern era, complete the system. Dates before the Christian era are preceded by a minus symbol. The UDC diverged very substantially here from Dewey, for the latter had introduced divisions by periods and sub-periods, by means of hierarchical decimal symbolization.

This system has the virtue of simplicity, but has two disadvantages: the ponderousness of the symbolism (nine signs are usually necessary, not counting the quotation marks, to express a period covering some years, for example, “1939/1945” for the period of the second world war), and the fact that, very frequently, approximate periods peculiar to each cultural area, even to each country, would be better adapted to the classification of historical phenomena. It is, in fact, in recognition of this factor that, in order to classify the ‘stages of development’ of languages, the UDC abandoned division by “ ” and adopted direct divisions, for example, 473 stages in the development of Latin: ancient Latin, low Latin, mediaeval Latin, humanistic, or 477 stages in the development of the Greek language. It seems rather inconsistent to have followed such a method merely in class 4, when it would have been doubtless most useful in class 9 history. It is true that conventions can always be established, similar to that for the designation of the mediaeval era, but they run the risk of varying according to the classifiers.

In class 7 fine arts, the UDC offers an alternative: it provides a classification by ‘styles’ which corresponds more or less to a classification by approximate eras (see 7.08), but adds that the ‘styles considered separately are indicated by means of common subdivisions of place and time’ (p. 268 of the abridged trilingual edition), which amounts to letting the user ‘shift for himself’ as best he can, and according to his individual preference.

Following the system adopted for ‘normal’ chronological divisions, inasmuch as we are now in 1960, the available notation remained practically unused after “196”. Rather ingeniously (although perhaps not too rationally) the UDC—which had not foreseen that it would still be used after the year 2999—made use conventionally of divisions “3/7” for ‘division of time according to various points of view’. Therefore, we have the following:

- “31” *past, present, future*
- “311” *past*
- “312” *present, now*

- “313” future
- “32” *seasons*
- “321” spring
- “322” summer
- “323” autumn
- “324” winter
- “33” *months*
- “339” weeks
- “34” *days and portions of days*
- “36” *time from the point of view of danger and need*
- “362” peacetime
- “364” wartime
- “37” *time from the point of view of professional occupation*
(time of service, of work, of rest)
- “4” *duration or period* (seconds, minutes, hours, days, weeks, months, etc.)
- “5” *periodicity*
- “6” *various eras* (non-Christian)
- “7” *manifestations of things in time*, phenomenology of time
(development, simultaneity, repetition, etc.)
- “71” evolution, development

Such, at least, is the scheme given in the trilingual abridged edition (p. 45). It will be observed that “4” practically duplicates “33/34” on the one hand, and “5” on the other. In his letter of April 1958, F. Donker Duyvis indicated “4” as corresponding to periodicity, and did not mention “5”, which would have certainly been more logical.

“71” seems to duplicate 111.62 movement, change, evolution—but since these general ideas are there regarded from the standpoint of ‘ontology, metaphysics’, “71” was probably considered more ‘neutral’.

Here, as in the common subdivisions of ‘place’, the UDC does not anticipate the use of the general relation symbol, and this explains such headings as “362” or “364” (it is true that, whereas there is a main number for ‘war’, 355, which could have supplied a composite number such as “3 : 355”, ‘peace’ does not itself appear in the UDC; there exist only special headings for peace treaties, demobilization of armies, etc.).

We have noted (p. 26) the existence of a type of ‘mixed’ spatio-temporal common divisions, under ‘places of the ancient world’. We also find an anomaly of the same kind in the ‘palethnological periods’, which are symbolized, not by quotation marks, but by the parentheses of place divisions, under 571 (prehistorical archaeology, division of . . . 57 biological sciences) in the number 571(11), which includes a rather large number of subdivisions (see p. 504 of the 1937 German edition) and which indicate clearly

that, for ill-defined periods, the direct division system is more convenient than the system of exact chronological limits which is a common rule in the UDC. Nevertheless, it represents a rather flagrant twist of logic.

Another 'twist of logic'—due to the same cause—concerns the geological eras. These are not located in the “ ” divisions, but directly in the main schedules, under 551.7 historical geology, stratigraphy (see the 1937 German edition, p. 483-4). And paleogeography is not expressed, as might have been expected had the UDC remained faithful to its principles, by a combination of divisions (...) and "...", but by other direct divisions, under 551.8 (divided, however, as 551.7).

To bring to a conclusion the question of time under the UDC, it can be observed that, whereas in principle three common divisions exist signifying 'past', 'present', and 'future', one finds in other places, expressed differently, the ideas of 'before', 'during', and 'after'—which, basically, are identical if the existence is implied of a 'point of reference' in time. Thus, among common subdivisions of standpoint by .00, one finds under .001.4 (tests, experiments) three direct divisions, 001.41/43 for: 'during setting up, trial runs', 'during use, working tests', and 'after alterations or reconditioning', which, theoretically, could equally as well have been realized by combination with "311", "312", and "313". Doubtless, the authors of the UDC wished here to make an economy of notation; this preoccupation seems often to be uppermost in their thinking and leads them frequently to prefer less 'logical' direct divisions to divisions by combinations of numbers, which are more 'logical', but less 'economical'.¹⁸

The order we have followed until now in examining the common notions in the UDC was somewhat naturally indicated by the order adopted by its authors in their arrangement of the 'auxiliary tables', which deal first with the signs of addition and extension (+ and /) representing the equivalent, roughly speaking, of the connexion symbol in logic, and then with common divisions introduced by special signs. We had set aside the language divisions by = and those of race by (=), believing that these notions were of a far less general character than those of space and time, which also have special UDC signs. A brief reference will now be made to them.

Subdivisions of *language* by = apply, in fact, solely to the language in which the documents are written. They do not therefore represent a characteristic of the *subject* of the documents, but merely an element pertaining to their *presentation*. They are purely and simply derived from class 4 linguistics, of the main tables, and

to a great extent are arbitrary; following Dewey, the UDC first set up a series of seven languages which, in 1876, at the library of Amherst College were considered the most important (English, German, French, Italian, Spanish, Latin, and Greek)—then the ‘other languages’, which in Dewey, were designated as ‘minor languages’. This represented a purely ‘practical’ order, answering the particular needs of a small American library of the last century. An attempt was later made at amending the order, to make it coincide a little better with current classifications made by linguists. The most important amendment consisted in grouping Greek and Latin under a ‘classical languages’ heading, in order to free 48 for ‘Slavic and Baltic languages’; by an extension of its content, 44 became ‘Romance languages’, French being 440; similarly 43 became ‘Germanic languages’, German being 430 (Dewey, as is known, did not use classification numbers of one or two figures; there was therefore no ‘collision’). But there remained, however, a 49 entitled ‘Oriental and African languages, and other languages’ which is, in effect, a kind of ‘catch-all’, in which 491 is entitled ‘miscellaneous Indo-European languages’, but does not include the Indo-European languages already classified in 42/48, where is found a 494/499 isolating and agglutinative languages, which is a class based on type and no longer on an historic point of view. The order and likewise the distribution of notation are not very international in character.¹⁹

Class 8, literature, is divided in the same way as class 4, and may be subject to the same criticism. The UDC corrected an anomaly in the DC, which had reserved a sub-class (81) for American literature: it changed it to 820(73).

The ‘races’ divisions by (= . . .) apply, themselves, to the subject of documents and no longer to their form. These also are purely and simply drawn from class 4 (with a few specific additions, such as (=1.5) races and colonial peoples, (=1.6) half-breeds, (=1.81) natives, aborigines, (=1.82) non-natives, creoles (?), (=1.100) cosmopolitans, (1=2) peoples of specified regions, divided as (2) physical place, (=1.3) peoples of specified countries in antiquity, divided as (3), (=1.4)/(=1.9) peoples of different countries (i.e. divided according to ‘modern’ political geography, as (4/9)).

The principal division of races is, however, in conformity with the (arbitrary) division of languages examined above, with a few modifications such as the denomination of (=9) which becomes ‘Oriental races in general, coloured races’, which produces the rather paradoxical result that the ‘Aryans’ (=91) are thereby classified under ‘coloured races’, and the addition of a (=2) for the ‘white race in general, Occidental races, Nordic races’. Evidently, such a ‘classification’ is beyond any criticism: it is

merely an arbitrary 'series of pigeonholes', and it is not necessary to dwell on it at any length.

There remain only, as *at present* forming a part of the UDC's common subdivisions, the 'common subdivisions of point of view' by .00, which we have already mentioned earlier (p. 26). There are fundamentally very few truly general ideas among them; they relate to 'operations' in the sense of Vickery (actions undertaken by man). They are arranged according to an order which corresponds roughly to a practical order for the filing of records of a business enterprise or an administrative organization: theoretical points of view; points of view of realization; economy and finance; utilization and operation; furniture, fittings and equipment; premises and establishments; personnel; organization; social and moral.

As we have observed (p. 23, Note 10), analytical subdivisions by hyphen, which were up to the present set aside especially for use in one or more classes, are now to be promoted to the dignity of 'common' subdivisions. The first to have been made the subject of a full-scale examination, with that object in mind, was -05 devoted to the 'person'. In this, one will perhaps notice the influence of Ranganathan's ideas on the 'personality' category, which we shall examine later on, but in the UDC the interpretation of the word is far more restricted (and, it must be added, much more consistent with the current and normal conception of the word). According to note F58-67 of 25 August 1958, the divisions of the said -05 are to be as follows:

- 052 according to rank, position in the service, remuneration
- 053 according to age
- 054 according to race (formerly subdivided by direct division, and now by the use of (= . . .))
- 055 according to sex
- 056 according to predispositions, constitution
- 057 according to profession
- 058 according to social class, or family situation
- 059 according to salary

These are derived from 616 pathology, for the needs of which they had formerly been created, and they still maintain certain traces thereof (for example -059, and, to a lesser degree, -056). They had later been 'transplanted' into 3-05 interested persons, in the social sciences, and 658.3-05 categories of persons, in personnel administration, division of industrial and commercial organization.

It is difficult to conceive the idea of a 'person' as a truly general notion, and we will not linger on -05, which, in fact, falls within more specialized categories than those to which, in principle, this study is devoted. It can be noted, however, that if a

class exists where the concept of the person (or better, personality) applies particularly well, it is that of psychology. But -05 is not used here. Under 159.9 psychology (this long number results from the cancellation by the UDC of the completely obsolete Dewey divisions under 151/159; in order to avoid any conflict between the Dewey classification identifications and those of the UDC, the latter does not use 151 to 159.8, and has created new divisions under the last identification, not used in the DC), there is 'comparative and differential psychology' under 159.92, where are located, under direct divisions, sexual psychology (159.922.1), racial (159.922.4), child (159.922.7)—together with, it is true, many other things which could not have been expressed by -05, such as phrenology and graphology, or biopsychology. Neither is -05 used in 331 work, employers and workers: a division by .051 relates to 'race, age, sex', and direct divisions 331.3 and 331.4 relate to the work of children and women. The profession is found (among other things) under 331.7 types of workers, qualified and non-qualified work. Similarly, under 362 charitable institutions, social work organizations (French title; the English title, social welfare, is less obsolete), divisions by the characteristics of the persons aided are obtained by direct division (for example, 362.7 provisions in favour of children, child welfare). Direct divisions also appear in 613 'private' hygiene, where professional hygiene is found under 613.6, and hygiene in relation to race, age, and sex under 613.9; even in medicine, from which the -05 divisions are derived, one finds an exception to their application for 618 gynaecology, which is a direct division—most probably because it constitutes a medical 'speciality'. On the other hand, -05 applies to 371 schools and educational systems, 'superadding' itself, however, to an older division 371.04 'Der Unterricht im Hinblick auf die Schüler' (incorrectly translated in the trilingual abridged edition as 'education by specified categories').

We have now concluded our observations on the 'general categories' as presented in the UDC. This general picture could perhaps be completed by searching for the general notions which may be more or less hidden in the details of the direct divisions and the divisions by - or .0. Such a search operation would exceed the limits of the present study, and we shall limit ourselves to essential points.

The *most general notions* are found in classes 11/12 metaphysics: 111.1 being, 111.3 substance, 111.4 accidents (11.42 qualities, faculties, aptitudes) 111.6 action (111.61 possibility, 111.62 movement, change, development), 111.82 unity, multiplicity, 111.83 truth, 111.84 goodness (Gutsein). Movement is

found again at 116;²⁰ matter is in 117, strength, energy in 118, quantity, number in 119. We have already seen space 114 and time 115. Under 12, 122 causality, 123 freedom and necessity, with chance in 123.11, as a subdivision of freedom. Under 124 teleology one finds .1 order (.12 contingency), .2 significance or meaning (Sinn), .3 aim, .4 *causa exemplaris* (.42 type, .43 ideal), .5 value, .6 destiny. The finite and infinite are in 125. 12 concludes at 128 and 129 which relate to the soul, life, and death, and 'origin and destiny of the individual soul'—which one would rather have seen included under 13 philosophy of the spirit.

The *logical categories* are found under 16, still within 'philosophy' in the Dewey sense, preserved by the UDC, following the occult sciences, philosophical systems and psychology. Logic finds itself inserted between this somewhat odd group and classes 17 and 18 moral science and aesthetics. It does not seem particularly necessary or useful to repeat here the details of its UDC classification, which in any case is not particularly developed and will be found conveniently accessible, on pages 63-4 of the trilingual abridged edition, practically in its entirety. It may be observed that 167 scientific research, duplicates .001.5 in the common subdivisions of point of view, as well as various subdivisions of 001 (science and knowledge in general) of the main schedules. On the other hand, the concept of scientific laws has particularly good luck: it is encountered three times, in 001.6 (scientific laws), in 167.6 (laws, rules) and in 113 (general laws of nature).²¹

Some of the general notions which we have just encountered, regarded from the 'philosophical' point of view, are also found in other locations of the classification, as was to be expected considering the general structure of the UDC to which we had already drawn attention (p. 22).

Action (111.6), is obviously present almost everywhere in the schedules, and it would be out of the question to note all its appearances. The same applies to qualities or properties (111.42). *Movement* (111.62 and 116) is also found in many places in the Engels sense, i.e., as 'mode of existence of matter'. In its more restricted meanings, it is found in 531/534 (mechanics), in 539.3 (mechanics of elastic solids) in 541.12 (chemical mechanics), in 541.124 (chemical dynamics, mechanism of reactions), in 521.1 and following divisions (celestial mechanics), in 551.2/3 (geodynamics) and in 441.465 (dynamics—and statics—in oceanography), etc. In various divisions of 621-1 (general) characteristics of machines, are found numbers for characteristics of movement and the transmission of movement to these machines.²² It would be an endless task to seek in the UDC tables all references to notions of *evolution* and *development*. The alphabetical tables (including all synonyms, quasi-synonyms and composites) would permit a

preliminary approximation, but would by no means be sufficient. In addition to its place in 122, *causality* bears another number in 'general principals of physics', 530.16, under the title 'principle of causality and probability, principle of indetermination'; this is where is found the Brownian movement. Near by, under 530.15, is found the 'principle of action and reaction', with a division assigned to resistance, inertia, hysteresis (530.152.1).

Number and *quantity* (119) are also found almost everywhere, of course primarily in 51 (mathematics). Then under subdivisions by -, such as 575-184 (number, in genetics) or 621-184 (characteristics of machines and apparatus according to the number), omitted in the trilingual abridged edition, probably due to the fact that the writers had not understood its meaning; F. Donker Duyvis clarified the above by indicating that this referred, for example, to the number of pistons in a rotating engine.²³ But, especially through the medium of *measurement*. The treatment of the latter in the UDC does not seem to be particularly satisfactory. Metrology is given in 389 which is a division of 38 commerce, communications, transport; this at first seems to be a somewhat unexpected place. It is not surprising, therefore, that there was little temptation to designate 'measure of . . .' by 389! We have, therefore, very generally, a number of direct divisions, of which a list follows of those we have been able to locate:

159.938	psychophysics, psychometry
159.98	psychotechnique
330.115	econometry
332.4/5	currency (under 332 finance)
338.5	prices (and value: part of 'production of wealth, economic condition')
338.971	study of economic fluctuations (Konjunktur-forschung)
522	practical astronomy (where are found the instruments, including the measuring instruments)
523.872	recording and measurement of stellar spectra
525.1	earth constants
525.4	geographical coordinates (determination on land)
527	navigational astronomy
528	geodesy, photogrammetry
529	chronology
531.7	measurement of fundamental sizes, geometric and mechanic (numerous divisions: see trilingual edition, p. 129; German edition of 1937, p. 377-9; we have not seen the German edition of 1958)
532.137	viscometers
532.217	measurement of levels
532.57	measurement of the speed of flow (hydrodynamics)
533.4	barometers

General classification systems

- 534.44 quantitative analysis of sound vibrations
- 534.6 acoustical measurements
- 534.8 applied acoustics (various subdivisions relate to measurement: .839, .844.1/2, .845.1/2, .88)
- 535.2 propagation and energetics of light radiation, photometry (the latter is found to be distributed among various divisions, mainly in .24, but also in .212.4, .221/223)
- 535.321 methods for the measurement of: transmission, reflection, refraction, absorption, emission (of light)
- 535.322 measuring instruments
- 535.568.1 polarimeters
- 535.568.2 saccharimeters
- 535.65 measurement of colours (but the measurement of wave-length of colours is found at 535.61.08)
- 535.853.225 spectrometers
- 535.853.26 prismatic spectrometers
- 535.853.4 interferometers (also at 622.412.3)
- 536.5 thermometry (includes also thermoregulation)
- 536.6 calorimetry
- 537.7 electrical measurements ('classify preferably under 631.317' advises the trilingual edition)
- 538.7 terrestrial magnetism (measurement)
- 541.141.9 actinometry
- 541.147.7 sensitometry
- 542.3 measurement of weights and volumes (in experimental chemistry)
- 543.279 gazo-volumetric analysis
- 543.86 radiometric analysis
- 545 quantitative analysis
- 550.93 measurement of geologic age
- 551.464.5 qualitative analysis of sea waters
- 615.14 dosage of medicine
- 621.317 industrial electrical measurements
- 621.386.82 radiochronometers
- 621.646.3 rheometers
- 621.693 pulsometers
- 621.753 gauging and gauge-making
- 622.412.3 interferometers (also at 535.853.4)
- 622.413.6 catathermometers
- 658.542.1 time study (in organization)
- 658.73 cost price
- 658.8.03 selling prices, tariffs, charges
- 681.1 wheel mechanisms, clockwork, counting and control devices (this rather vague heading covers all types of measuring apparatus: clockwork, time meters, consumption meters, etc.—with all the calculating machines)
- 681.2 instrumentation, instrument making, measuring apparatus

681.88	directional listening and sound measuring, listening apparatus
771.53	sensitometry (also at 541.147.7; no cross reference in the trilingual edition)
778.38	photogrammetry (with a cross reference to 526.918)

It is very probable that we have omitted a great number of other numbers hidden among the thousands of pages of the various editions of the UDC.

Furthermore, numerous divisions are devoted to measurements or measuring instruments which are obtained by other processes than direct division: (a) By use of the symbol of general relationship : - this is a rather rare procedure, which we have not often encountered outside colloidal chemistry, 541.18:531.7 (and its divisions) and mining topography, 622.1:526; it should be noted, also, that the trilingual abridged edition makes no reference to this particularity; (b) By analytical subdivisions by means of the hyphen, for example, 616 pathology, where are found numerous instruments and measurement methods under -073, divided subsequently 'as 53 physics'; many of the headings tabulated above are, in effect, repeated, and a few others which are not, such as -073.26 which contains stalagmometry. Similarly in 615 pharmacy, therapeutics, is found an analytical division -015.3 dosage including dosimetry—which seems also to duplicate 615.14—and another in -092.22 for the measurement of activity in pharmacological physiology.

Nevertheless, the most common procedure and, one might say, the 'normal' one seems to be the use of analytical divisions by .0—but there is no uniformity in the figures which follow. For means of transport (except vehicles on rails), in 629.1, .05 is used; in 551.46 oceanography and 551.48 hydrography of superficial waters, .018 is used; for 539.1 nuclear, atomic, and molecular physics, .074 is used (meters and detectors; memorandum PE 663 of 2 June 1959). Most frequently, however, .08 is preferred: such as under 52 astronomy (but here lies a danger of 'collision' with direct divisions which, as we have seen, are very numerous in this class); under 53 physics (same observation); under 550.34/38 geophysics; under 551.5 meteorology (here the point is omitted, one does not write 551.5.08, but 551.508; on the other hand, a recent memorandum PE 657, of 17 December 1958, introduced a 551.501.71 division for 'methods of observation and calculation of composition and density', and analogous divisions in .721/777 for other phenomena); under 620.1 testing of materials; 621.3 industrial electricity, electrotechnics (it is difficult here to understand how a distinction is to be made from the direct division 621.317 industrial electrical measurements; no indication in

the tables permits us to determine this and the choice is probably left to the individual user, with all the attendant risks of confusion). In classes 578 biological technique, 581 general botany, 591 analytical zoology, 611 anatomy, 612 physiology, and 619 veterinary art, .087 biometry is used.

In the aggregate, this picture of the treatment of measurement and measuring instruments in the UDC does not appear very satisfactory.²⁴ It gives the impression of developments carried out in a disorderly fashion by successive uncoordinated additions. Most of the existing direct divisions have been retained; wherever no provision had been made for measurements, such and such a method has been chosen according to the needs and also, doubtless, according to the inspiration of the various specialists who were engaged in revising 'their' UDC chapter.

Chance or hazards, which we encountered in 123.11, is—under a different aspect—in 519 combinative analysis, computation of probabilities, etc. (this 'etc.' includes, among others, the theory of aggregates). But 'statistical science, theory and methods' is found under 311, a division of a statistical class included in 3 social sciences. The trilingual abridged edition does not indicate how statistical methods in other sciences should be classified; only the English alphabetical index provides a cross reference 'Statistics 31 cf. Biometry'; 578.087 is, in fact, the analytical subdivision devoted to biometry in biological technique; following an examination of various cross references or explanations (not all to be found in the trilingual edition) it will be ascertained that the same .087 analytical subdivision applies also in the case of 581 general botany, 591 analytical zoology, 611 anatomy, 612 physiology, and 619 veterinary science.²⁵

As we have seen, probability is found, in physics, under 530.16.

We will conclude this inventory of general categories in the UDC by *control* and regulation. A detailed analysis of these notions would take us outside our bounds: their relationships with ideas of norm, law, government, and those of freedom, finality, would better find their place in a study of a philosophical character—or semantic, in the sense of Trier or of Matoré. We are all aware of the present development of this entire field of control, regulation, servo-systems, phenomena which form the basis of automation, and are studied by cybernetics, the theory of control and of information.

Under cybernetics, in the French index of the trilingual abridged edition, is found only the number 621.391, which is entitled in the systematic table, under telecommunications, 'generalities, cybernetics, and theory of information in their relationship with electrical telecommunication'; similarly in the

German index, at 'Kybernetik'. On the other hand, the English index does not mention, under cybernetics, this 621.391, but gives a mere cross reference: 'see automatic control'. There, one finds 621.3.078 (which is not in the main schedule, where the subdivision ceases at 621.3.07 regulation of electrical and magnetic quantities) and 621.316.7 regulation, regulators, under 621.316 distribution and control, itself a part of 621.31 electrical power engineering—but especially the analytical divisions by hyphen 621-52/-55 which, in fact, relate to servo-mechanisms.

PE note 656 of 8 November 1958, develops, in another connexion, divisions -50 and -551, following a suggestion made by the International UDC Panel of FID for Measurement and Control. It is possible, therefore, that this -5 is more or less destined, in the future, to develop as a common subdivision for control—perhaps also, even, for information in general, which does not have any UDC 'headquarters' ('information-technique' remained blank in the French index of the trilingual edition; commercial information services are found at 659.21, the Ministry of Information at 354.36 . . .), for, in the said PE note there is, under -501, 'principles and theory of signal transmission'. But what an undertaking for the UDC Panel should it have to set in order all of the existing divisions for regulation, control, etc. Probably, however, what occurred in connexion with measurements would take place here, i.e. a coexistence of older numbers with direct division or .0 . . . (a certain number of these will be recognized when consulting the various indexes of the trilingual edition, under control, rule, regulator, etc., and under their English or German equivalents; it is always advisable for the searcher not to limit himself to one index only, as we have seen), and new numbers divided by -5.

What conclusion can be reached from this long *exposé*—which we have developed in more or less detail, recognizing that the UDC is at present the sole 'norm' (or semi-norm) existing on an international scale in the field of classification?

According to the individual state of mind (or of the same person at different times), one can either admire the richness of the UDC's means of expression—let us say even their superabundance—or deplore the risks of confusion between different numbers with a very similar meaning, or again criticize the inconsistency of many of the methods. The UDC is probably the only classification which can 'tell all'; among the mass of detail in its tables, it is difficult not to find a 'niche' for each subject, provided it does not relate to a development of too recent a date to have gone through the rather slow process of extension

proposals. But it has become a nomenclature more than a classification, and this was unavoidable for an enumerative type of classification, such as was the Dewey classification originally; on the basis established by Dewey have been added 'superstructures' in the shape of common subdivisions, by colon, hyphen, etc., all of which, however, has not transformed the system.

It is understood, of course, that the latter can evolve, develop, and to a certain extent be rationalized—this has been proved, thanks to such eminent men, with their encyclopaedic knowledge, as Otlet and La Fontaine in the past, and F. Donker Duyvis today. Such development and standardization, however, bear certain limits, due, on the one hand, to the basic principles upon which the Decimal Classification has been established, and, on the other hand, to the fact that it is a standard, or semi-standard; therefore account must be taken of certain needs or, at least, wishes of users for stability.

As written by Donker Duyvis (note F58-75 of 4 September 1958), 'it is always difficult to broaden and generalize the scope of existing divisions, whereas any fool can narrow the meanings of existing divisions by subdivisions'. This is all the more difficult when starting from a system such as Dewey's, conceived for particular objectives (the classification of a small American college library, at the end of the third part of the last century) and on the basis of conceptions which are no longer those of today: the existence of 'fields' of knowledge sharply separated; the separation of science from its applications, etc.

It is not possible to 'renovate' the UDC beyond a certain point—otherwise it must be transformed into another system. It will remain, due to the force of circumstances, a classification of the 'strong hierarchical type' of Mooers—and it can be considered that this type of classification is no longer adapted to present requirements of documentation, and, in particular, fails to meet the needs of automatic information retrieval. This does not imply, of course, that it will rapidly disappear: it has become an 'institution', and institutions, following the rule of Spinoza, tend to persist in their nature. There are, and there will continue to be, users of the UDC; their number will doubtless increase. It is probable, however, that the future lies elsewhere—and it is necessary to start preparing it now.

12 From the UDC to the Colon Classification system: the classification systems of the Library of Congress and of Bliss

We shall cover much more briefly the classification systems which originated during the period between 1895 and 1924, the latter

date being that of the beginning of the elaboration of the Colon Classification by Ranganathan. That period, in fact, offers no innovation of an importance equal to that realized by Otlet, at least from the standpoint which interests us here.

Starting in 1901, the classification system of the Library of Congress (LC) began to appear; based, incidentally, upon that of Cutter. Even more than the UDC, this is a system of traditional type; it is entirely enumerative, and it would be useless to seek therein an expression of general categories: specific subjects only will be found. Fundamentally, it is, after an interval of two centuries and a half, an undertaking very similar to that of the great librarians of the Bibliothèque Mazarine, and later those of the Royal Library in Paris, i.e. a careful and detailed arrangement of the greatest collection of books in the world, on the basis of groupings of the latter, let us say by affinities. The LC librarians are therefore in this way the direct continuators of Gabriel Naudé and Nicolas Clément.

In 1910, Henry Evelyn Bliss published in the *Library Journal* 'A modern classification of libraries, with simple notation, mnemonics, and alternatives'. The 'Bibliographic Classification' (BC), as it was later named, did not, however, appear in abridged form until 1935, after publication by the author of two very interesting volumes on theory, the first relating to the classification of science (1929) and the second concerning library and bibliographic classification (1934), and the complete edition did not begin to appear until 1940. Bliss developed the use of auxiliary tables (systematic auxiliary schedules), less fixed, however, than those of the UDC: in fact he made a distinction between 'constant' and 'adaptable' mnemonic schedules. His principal effort was directed to one point: to draw bibliographic classification nearer to what he termed the 'scientific and education consensus'—the scientific and pedagogic order of subjects of study.²⁶ His work, from this standpoint, has historical importance; it represents a reaction against the purely pragmatic, practical aspect of the Dewey or Cutter classifications—or even that of Ranganathan, with his 'canonical divisions'. He recognized also, in a certain measure, the inadequate character of a classification system in which the main classes are based on disciplines and not on subjects studied, and established numerous 'alternative locations', so that the same complex science could be placed in several principal classes. But he did not follow the avenue opened by Otlet for the introduction in bibliographical classification of relationships other than a hierarchical relationship, and more so even than the UDC, the BC remains a purely enumerative system of a traditional type. It was only in 1948, perhaps under the influence of Ranganathan's ideas, that he advocated the use of relationship signs.²⁷

It may well be that the most truly *new* system during this period was J. Kaiser's *Systematic Indexing*, published in London, by Pitman in 1911, the most important application of which was made at the Nobel d'Ardeer manufacturing plant, in England, now a part of the Imperial Chemical Industries group.²⁸ Kaiser had in fact systematically separated, in an alphabetical subject catalogue (containing cross references to 'higher collectives' or to 'lower specifics'), what he described as 'concretes', substances or equipment materials, specified, moreover, by qualifying statements serving as attributes, and by 'processes', actions carried out or suffered by the 'concretes'. This represents, probably, the first example of uniform application, within the entire scope of an information retrieval system, of 'categories' of terms; Kaiser seems, therefore, to have been a precursor (unknown to them) of Leroy and Braffort in France, and of Cherenin in Russia. It is also possible that he may have inspired Ranganathan, although the latter makes no mention of Kaiser.

13 The Colon Classification

Shiyali Ramamrita Ranganathan began in 1924, in London, the elaboration of his classification scheme, the first edition of the latter appearing in Madras in 1933 and volume 1 only of the 5th edition (to our knowledge) having been published in 1957. Inasmuch as the Colon Classification (CC) is still in a state of evolution, it is somewhat difficult to examine it in detail, more especially since volume 1 of his first edition contains merely abbreviated schedules, leaving to separate sections of volume 2 the elaboration of the details of the subdivisions.

The CC has remained essentially, up to the present, a library classification system; it does not seem to have been adopted extensively for information retrieval systems 'in depth'. Volume 1 of the 1957 edition is, in actual fact, expressly intended for the classification of *books* (cf. preface).

Under the term 'relations', the CC—remaining faithful to its method of using words in a particular sense and creating a kind of 'Ranganathanian language', which certainly does not facilitate the comprehension and use of the classification system by other than initiates—envisages, by naming them 'phase relations' or 'intra-facet relations' (according to whether they are to be applied to classification numbers or numbers forming a part of different 'main classes'²⁹ or of the same main class), five types of relations to which he assigns a specific symbol (lower-case letter). These are: 'general relation' (*a* as phase relation, *j* as intra-facet), identical in principle with the colon in the UDC, but actually of less general application owing to the existence of other types of

relations which tend to restrict in the CC the application of this 'residual' relation;³⁰ the 'bias' relation (*b* or *k*), which relates to a subject stated in a manner especially intended for specialists in another subject (for example, 'psychology for doctors'); comparison (*c* or *m*); difference (*d* or *n*); and influence (*g* or *r*). As observed by Vickery,³¹ this list should definitely be completed.

In practice, however, the 'phases' or 'intra-facet relations' are not very frequently used in the CC, because—as Vickery so observes (*Classification and Indexing in Science*,* p. 41)—the latter, by its 'analysis by facets' and use of (fundamental) 'categories', already expresses by other methods a great number of relationships other than of genus to species: between a thing and its parts, its constituent substances, its properties, and the processes to which it can be submitted.

The fundamental basis of the CC remains, however, as in the case of the UDC, the hierarchical or inclusion relationship. The entire universe of knowledge and activity is divided into a certain number of 'main classes', at present numbering thirty-one.³²

These main classes in their turn are divided into a certain number of sub-classes (generally around ten, due to the fact that the CC, after enumeration of the main classes by means of Roman capital letters or Greek letters, uses a decimal notation of the Dewey type) which are obtained by the application of either 'canonical divisions', or a category 'personality'.

Ranganathan does not seem to have defined very exactly what he meant by *canonical classes*. Practically, these are 'traditional' divisions into specialties of a field of study constituted by academic tradition. For instance: mathematics (B) is divided into arithmetic, algebra, analysis, 'other methods', trigonometry, geometry, mechanics, 'physico-mathematics', and astronomy (the first five and the eighth of these divisions are themselves subdivided into 'canonical divisions'); physics (C) is divided into 'fundamentals', properties of substance, sound, heat, light and radiation, electricity, magnetism, cosmic hypotheses and 'special' (atom, nucleus, high pressure, low temperature, etc.); geology (H) is divided into mineralogy, petrology, structural geology, dynamic geology, stratigraphy, paleontology, economic geology, cosmic hypotheses; useful arts (M) is divided into production and description of books, journalism, domestic science (including also carpentry and the glass industry), textiles, clothing, masonry, photography, leather industry, packing and crating (other industries to be specified by the 'subject device', i.e., division by means of numbers borrowed from other classes of the classifica-

* Hereinafter abbreviated as *CIS*.

tion scheme—the equivalent of Dewey's 'divide as . . .', and, to a certain extent, of the UDC's division by :), physical education, sports, and games.³³ Fine arts (N) are divided into among architecture, sculpture, graphic arts, painting, music, theatre; philosophy is divided into logic, epistemology, metaphysics, moral science and esthetics, followed by three classes for philosophical systems.

The divisions obtained by application of the 'personality' category (of which the distinguishing sign is a comma, which can, however, be omitted when relating to divisions applying directly to main classes) have also, fundamentally, a traditional character. Ranganathan has defined clearly what he means by 'personalities'; his FID/CA report (document F (Comm) 59-48 of 15 June 1959) contains a paragraph (55) entitled 'Ineffability of Personality'. Vickery (*CIS*, p. 176) has identified it with the Aristotelian 'substance' and he attempted very strenuously to assign to it a concrete content. It would seem to us much simpler to make it the equivalent of 'subjects of study of a specified discipline, as these commonly serve as a basis for the traditional division of this discipline, or for whatever division the CC considers the more practical'.

Considered as such, 'personality' has no theoretical value; it is simply a label placed on one or more characteristics selected more or less arbitrarily to form hierarchical divisions of the highest order in a classification which seeks to express different 'points of view', but acknowledges as indispensable the existence of a fixed and immutable order among the latter. An examination of the divisions included under category [P] in various main classes of the CC reveals that one should not seek here anything more than an *a posteriori* standardization of a purely practical method.

The P divisions of library science (class 2) are categories of libraries, enumerated as follows:

- 1 'trans-local' (worldwide, national, etc.)
- 2 local
- 3 academic (schools, universities, and research)
- 4 business (i.e. firms and public administrative organizations, also religious organizations 4Q, Q indicating religion)
- 5 subscription (paying)
- 6 special classes (for children, prisons, hospitals, the blind, etc.)
- 7 private
- 95 'contact' (?)

The divisions of geometry (B6) are as follows:

- 1 line
- 2 plane

The Colon Classification

- 3 three dimensions
- 4 four dimensions
- 5 five dimensions
- 7 n dimensions

Those of astronomy (B9):

- 1 earth
- 2 moon
- 3 sun
- 4 planet (of the solar system)
- 5 meteor and comet
- 6 star
- 7 planetary system

Those of 'light, radiation' (C5):

- 1 visible light (divided according to the 7 main colours)
- 2 ultraviolet
- 3 X
- 4 gamma
- 6 infra-red
- 7 Hertzian waves
- 91 source
- 911 monochromatic

Those of engineering (D):

- 1 civil engineering (with no further division)
- 2 irrigation and drainage work
- 3 building
- 4 transport—track
- 5 transport—vehicles
- 6 mechanical engineering (divided into 'principle of mechanism and special device', hydraulics, pneumatics, heat, electronics, electricity)
- 7 nuclear engineering
- 8 municipal (sanitary) engineering

Those of agriculture (J):

- 1 horticulture
- 2 feed
- 3 food
- 4 stimulant
- 5 oil
- 6 drug
- 7 fabric
- 8 dye, tan
- 91 adhesive

Those of psychology (S):

- 1 child
- 2 adolescent
- 3 post-adolescent
- 4 vocational
- 5 sex
- 6 abnormal (genius, insane, blind, etc.)
- 7 race
- 8 social
- 9 animal (genetic)

Those of chemistry (E):

- 1 inorganic substance
- 2 hydroxyl, basic oxide
- 3 acid, acidic oxide
- 4 salt
- 5 organic substance
- 6 aliphatic compound
- 7 aromatic compound
- 8 heterocyclic compound
- 9 bio-substance

apply also for class F technology, with a few additions (division of 5 into plastics, food, alcohol, fuel, medicine, textile, dyeing, etc.; subdivision and modification of 9, the 95 of E pigment, becoming for example cosmetics). They are also applicable in mineralogy (H1), in economic geology (H7) and in mining technology, η , with the addition of a heading for gems.

Divisions [P] of botany (class 1) and of zoology (K) are derived from the traditional Linnaean classifications; those of breeding (λ) are similar to those for agriculture, with three modifications: those of medicine (L) are by organs, apparatus and systems; within classes Δ 'spiritual and mystical experience' and Q religion, the various religions; similarly under R6, Indian philosophy, the divisions are the various schools. In literature (O) and linguistics (P), they are the various languages; in the N divisions, fine arts, they are the styles: '... the style', writes Ranganathan (p. 1.116) 'is individualized by the country and the century of its origin. This means that the style is a combined manifestation of Personality at two levels.' Divisions by styles are, in effect, constituted first by an indication of the country of origin, according to geographical classification, mentioned later, followed by the century, according to the schedule of common chronological divisions of the CC.

The 'personalities' of geography (U) are the traditional divisions into: geography, mathematics, physics (including meteoro-

logy), human biology, politics, economics, and travel. Those of history (V) are the countries, derived from the common geographical subdivisions; those of political science (W) are the types of government, with a few other subjects of study such as Utopia, 'change in the form of a state' (revolution, passive resistance), nationality. Those of economics are: communications, transport, commerce, credit, public finances, insurance, industry; under sociology (Y) they are the social groupings; finally, under law (Z), they are the 'communities', i.e., nations or other 'localized' communities (by geographical subdivision), or 'other cultural groups'.

It appears useless for us to seek a 'general category' with some real existence which would allow us to give a common rational basis to these various series of divisions; these can only be considered as *enumerative* classes, for the most part arbitrary, such as had been formed for a purely practical purpose by older library classifications.

'Personality' appears at subsequent levels of the CC. After the first divisions of a main class, or of one of its sub-classes, have been obtained by means of divisions by 'personalities', these are frequently subdivided according to 'facets' derived from the 'energy' category (see below) and then by a 'second round' of 'personality' [2P]; there may be [P] divisions of second or third order without a previous 'first round' of energy [E], but, in such a case, they are given the name of 'second (or third) level personality' [P2] or [P3]—this is a rule, as a matter of fact, which is presented by the CC as though it corresponded to a law of nature, i.e. that one 'round of manifestation' of the personality category or the substance category cannot 'manifest' itself until after an energetic 'round' (Rules, p. 1.63—energy is therefore a privileged category 'whose manifestation alone can initiate Rounds of Manifestation' (ibid., p. 1.64). All of the foregoing merely reflects simple conventions, established with a special, but rather artificial, terminology, with a view to arranging the rather complex classification numbers in a specified order.

[P2] (second level personality) under construction (D3) relates to 'parts' of a construction; the same applies to all the D divisions, except D6 mechanical engineering, where [P2] applies to a series of divisions named 'field of application': generation, transmission, transformation, traction, lighting, machines; under D66 industrial electricity, traction is replaced by transport. Under chemistry, [P2] relates to 'combinations': halogen derivative; hydroxyl, alcohol, phenol; organic acid; nitrogen compound; sulphur compound; compounded with another element. Under mining technology, it is the 'workings': well, gallery, mine hoist, transport on the floor of the mine, surface transport. Under botany, [P2] is the organ, as in agriculture and in zoology; under medicine

[2P]—following a division [E]—is the ‘cause’ (when the [E] division is illness) or the ‘substance’ (when [E] is pharmacognosy). Under architecture, style already being [P1] and [P2] combined, one finds [P3] for ‘use’, i.e., the eventual purpose of the building, and a [P4] for the parts thereof; under sculpture and painting, [P3] is the object represented, but one finds in the series of these divisions a 7 numismatics. [P3] in music, groups at the same time the divisions for the type of music (dramatic, orchestral), and the subjects such as the lyrics (libretto), the notation. Under literature [P2] is the literary genre; in linguistics, it is dialect, slang, jargon; there is a [P3] for the elements: sound, syllable, word, phraseology, etc., and punctuation. Under history, [P2] relates to parts: the chief of State, the executive, the legislature, the party, the local authority, the judiciary. Under law, [P2] simply designated by facet ‘Law 1’ applies to the person, property, contract, crime, conflict of laws, procedure, courts, sources of law, sanctions a veritable ‘hotch-potch’. There is a [P3] and a [P4], and for certain divisions [P2] (see schedules, p. 2.124 and 2.125).

The second ‘general’ CC category relates to *energy*, introduced the numbers by a colon (:). The energy facet, state the rules, p. 1.62, ‘consists generally of problems, action, etc.’.

Under mathematics, one most frequently finds methods under this ‘hat’; in mechanics, the fundamental principles, movement, equilibrium, vibration; in astronomy, ‘problems’: chronology, geodesy, spheric astronomy, physics, theory, cosmogony. ‘Problems’ also in physics: such as for C5 (light, radiation), propagation, dispersion, interference, etc., and nature of light. ‘Problems’ again for engineering: applied mechanics, resistance of materials, drawing, estimates, execution, measurement. And in chemistry: physical chemistry, analytical chemistry, synthesis, etc.

Under biology, botany, and zoology, the ‘facet’ [E] relates to nomenclature, microscopy, as well as morphology, physiology, pathology, ecology, genetics, ontogeny and handling. Under mines, the preliminary work, preparation of minerals, dangers and accidents, sanitary measures. Under medicine, one finds under [E] ‘preliminaries’ (nomenclature, hospital, anesthesia), then morphology, physiology, and pathology, hygiene, ‘pharmacognosy’, ontogeny—this scheme bears some resemblance to that of biology.

Under the class entitled ‘mystical and spiritual experience’, [E] relates to nomenclature, ‘technique’, occultism, prophecy; under painting, one finds under this chapter, composition, colour, water-colour, oil painting... preservation. Under linguistics, the ‘energy’ relates to sound (which we encountered earlier in [P3]), structure, function, sense, composition. Under religion, [E] covers mythology, the scriptures, theology, religious practices, institutions, sects, heresies.

Under psychology, [E] is divided as follows: nervous reaction; sensation; 'characteristics of consciousness' (attention, abstraction, habits, etc.); cognition, conception; sentiment, emotion, affection; conation, movement; personality, character; metapsychology (which includes sleep, and dreams; the sub-conscious).

Under history, [E] relates to: politics, constitution, function, relationship of the State with the special classes of the population, its relations with the citizens, sources, archaeology and the other 'related sciences', archives. Under economics it groups consumption, production, distribution, transport, commerce (which is already in [P]), value, organization. Under sociology, the grouping is extremely diversified; physical characteristics, activity, pathology, population, personality, equipment.

There is no 'energy' under literature, nor in law—at least in so far as the CC is concerned.

It may be concluded from this enumeration that here again the concept is vague, conceived after the event in order to cover divisions of a very varied character; one does find, however, mainly 'processes' and 'operations' with a few 'properties'.

As to energy in the usual sense of the word, it is found in physics, under C14, C82, C84—three numbers, hardly distinguishable, of the 'personality' facet—and its different forms are also included in the [P] category of canonical divisions of physics, and of class C engineering, etc.

Ranganathan's third general category is that of *matter*, indicated in the notation numbers by a semi-colon (;). The 'matter' facet, as stated by the Rules (p. 1.62) 'usually consists of materials used for construction, consumption, etc.' but they add that 'in some subjects, ranging from C physics to M useful arts, it will be seen that matter manifests itself as personality'—and, in fact, we have encountered much 'matter' under [P], [P2], etc.

In the CC's present edition, the [M] facets appear only in D engineering, (p. 1.91 of the Rules—but the schedules, p. 2.41 to 2.46, do not provide any corresponding division), and under three divisions of fine arts: sculpture (ND), painting (NQ) and music (NR). The [M] divisions of sculpture and painting are more or less identical; they relate to the material used for the sculpture, or on which the painting is painted (but not the nature of the pigment, which is found under [E]); for music, they relate to the instruments.³⁴

The general concept of matter, in its usual sense, is found in C13, in the fundamentals of physics, and in two numbers of C8 (cosmic hypotheses): a first time linked with energy (C82), and the second time alone.

Now, therefore, the three categories [P], [E], and [M]

appear to us in reality as kinds of 'false windows', aesthetic arrangements, perhaps, but they do not correspond to any tangible concrete reality. It could not be otherwise with categories as general, as vague, and as little exclusive of each other.

There remain, under the five categories, those of *space* and *time*, which bear the same distinctive symbol: the point or period (.)

It would be a serious mistake to believe that in the CC, notions of space and time are found only in the divisions of the [S] (space) and [T] (time) categories, introduced by this period.

In so far as *space* is concerned, the question relates solely to *geographical* divisions (this is, in effect, the title of chapter 4 of the schedules, which furnishes the details)—and not place more generally, as in the UDC. They present an advantage over the latter, i.e., that by applying the 'subject device' it is possible to specify very easily such divisions as 'Moslem countries', .1(Q7), Q7 being the symbol for Islam in the religion class. These are political geography divisions; the physiographic divisions (which will be given in detail in volume 2 of the fifth edition) appear as subdivisions at a second level.³⁵

Space, properly speaking, appears in B6 geometry; in C15, under fundamentals of physics (with three divisions: linear dimension, surface, volume); in C85 under cosmic hypotheses; the conception of space is in S42 under psychology (with that of time and movement). There is a division entitled 'dimension and shape' by :621 under physical astronomy, and another :635 for surface. Space, from the mystical standpoint, is in [P2] of Δ , division 5.

The environment in ecology is found under G:55 (biology), I:55 (botany) and K:55 (zoology); ecology in agriculture is J:95; on the other hand, 'ecological groups' are found in G95, I95, and K995 (first round of personality). There is a cosmic ecology in B95. Under Y sociology, one finds only one division, 75 ecological method, in [2E].

As in the UDC, *morphological concepts* (form, structure) are much dispersed: some are found in B621, already mentioned; under [P2] of D411 and D415 (highways, railways, e.g. D411,5 curve); a division :11 for structure is found in the first round [E] of D; molecular structure is at E:218 (first round [E]) and structure of crystals is H1:82. Structural geology is a canonical division, H3. Morphology is at G:2 under biology, I:2 in botany, J:92 in agriculture, K:2 in zoology, λ :2 in breeding or raising, L:2 in medicine; P:2 in linguistics. Musical form is located in division 2 of [P3] under music (NR); Gestalt in psychology appears only as a 'system', i.e. a particular theory the number of which is constructed with the chronological division corresponding to its date

of origin, SN. Geomorphology is U21 (first round of personality). Notions of form are found under Y:2 in sociology. It will be observed that these notions tend to regroup under a division :2 (first round of energy), but it is merely a tendency, with numerous variations and exceptions.

The CC index does not indicate any 'deformations', but they are found in C2:5 (elasticity, under the properties of matter). 'Form-work' is in D-76.

On the whole, the CC vocabulary is definitely not as rich in morphological divisions as that of the UDC—subject to what may appear in volume 2.

As in the case of space, *time* does not only appear in the 'chronological divisions' which are indicated on p. 2.6 of the schedules, and which have been indicated by a point followed by an upper case letter, in turn followed by decimal figures. The latter only serve the purpose of specifying the periods of the history of a subject (see rules, p. 1.62, especially the observation under 652). But, without the point, chronological divisions are also used to construct numerous numbers for divisions listed under 'personality' or under other 'categories',³⁶ of which they then adopt the particular symbol (comma, semi-colon, or colon).

Chronological divisions are, in fact, used and even abused by the CC. Medical 'systems' are also specified (e.g. LL, homeopathy), artistic styles (by combination with geographical divisions, e.g., NA5,D Romanesque architecture—which, as a result, comes before Etruscan architecture NA52,C; NA5 being European architecture and NA52 Italian architecture!), literary periods (following designation of the literature by applying language divisions, then literary form divisions), the successive stages of a same language (e.g. P122,A ancient French), Christian churches not specified by direct divisions of Q6 and 'other religions' Q8, educational systems (e.g. TN1 Montessori schools). But also available by 'chronological device' are: proportional representation (91M under 'systems' of history . . .), socialism (in general, under X:3M, under the facet [E] of economics, distribution division; syndicalism, on the other hand, is found under XN16 and Communism under XN17), joint stock companies (XM5) and trusts (XM8—specified as 'American'). This 'chronological device' is also used to designate many other things: details of functions, equations, etc. under mathematics (see B13, B23, etc.), theories on the nature of light (C5:8), Compton effects (C5:38M2), Raman (C538N28) etc., wave mechanics (CN2), psychoanalysis (SM9), intelligence tests (T:51N).

It is stated (Rules, p. 1.48) that 'the variety of uses to which the Time Number is being put is one of the elements that has invested the Colon Classification with a great degree of hospitality'.

It is, indeed, a convenient method for finding a location for any notion which one finds it difficult to classify logically in a hierarchical classification or which had not been foreseen at the start in the divisions. But it cannot be considered other than a makeshift—an expedient.

Time, in general, is under C16 fundamental physics, then twice at C8 cosmic hypotheses: C81 time, space and C86—time alone. It is found again at [P2] of Δ , under 6, from the standpoint of 'mystical experience'. It is under S:42 conception of time, space, and movement under psychology. As in the UDC, measurement of time is found under astronomy, B9:1.

The seasons are classed in ecology, under G:556, and biological ages generally also under a facet [E]—under divisions G:75 to G:78, in ontogeny; the same divisions are applicable in botany (I) and in zoology (K), and those of ontology are applicable in medicine (L), but there 'child' is also found again in L9C and 'old age' in L9E, where they appear under the heading 'Specials'. In psychology, the ages are in S1 to S3 (first round of personality); under education, T1 to T3 correspond roughly to divisions by age (but there is also in T3 education of adults, a T35 for foreigners, and a T38 illiterates). In sociology, Y1 is 'by age and sex', then under Y:34 ceremonial, are found divisions Y:341 to 344 relating to age.

Geological eras appear twice: once in the [T] 'chronological divisions' under A 'before 9999 BC', with five decimal divisions from 1 to 5, from eozoic to quaternary, and a second time under H5 stratigraphy, with the same divisions and a few subdivisions. The prehistorical eras are, in sociology, under Y71 (division of 'race as a social group'): tertiary is at 7114 and quaternary at 7115, then comes the stone age 714, and the metal age 718—followed by heading number 7196 for 'lost race'.

Notions of 'pre-' and 'post-' appear rarely: 'preliminary operation' can, however, be noted in the treatment of minerals η :21, as well as 'pre-secondary' T, and 'post-graduate' T45, under education. The length of time notion is found under [P4] of law, for injunctions by duration 56, and in B9:194 length of the day; longevity is also found in S791 division of personality, ego, character, under psychology.

Much as in the UDC, the CC contains more or less common divisions of *language* (schedules, p. 2.26 and 2.27), which serve in linguistics and in literature, and also—in imitation of the UDC—for the majority of divisions of *race* (see Y73(P1) to (P41); schedules, p. 2.117). We examined them cursorily in the report mentioned in Note 12. These divisions of *language* recognize only

three families of languages (and, by the same token, of races), Indo-European, Semitic, Dravidian; the entire remainder is classified by geographical order, e.g. 4 other Asian languages, 5 other European languages, etc.

There exists in the CC no division in which one might find a group of *very general concepts*, such as those encountered in classes 11/12 of the UDC. CC's metaphysics division, under R3, as a matter of fact, is devoted from R31 to R55 to a kind of classification by types of philosophical systems (which are also the object in R6 and R7 of two canonical divisions for the 'favoured system (1)'—the Indian philosophy—and the 'favoured system (2)', as well as of divisions by country in R8 for the 'other systems'). However, in this same R3, are found R36 teleology, and R37 value. These two terms are encountered again under [P2] or R4 moral science, divisions 6 and 7. The idea of cause, as such, does not exist in the CC; finite and infinite appear in composite terms under mathematics; group is in B27, under algebra, but aggregate is completely missing. *Logical notions* have been made the object of a very brief enumeration under R1 and R2. 'Law' exists merely in the juridical sense, and in physical chemistry, under E:211 law of chemical combination. 'Order' is found in library science and in law (Z, [P2] 9597, declaratory order).

It would seem to serve no useful purpose to determine here how CC treats notions of *action, movement, evolution*, etc. It acknowledges them only in specialized 'contexts'—and classifies them generally under facet [E]—not, however, without some exceptions, for instance, the action of water, of air, etc., in external dynamic geology is found in the canonical divisions H422 to H428.

As to *number and quantity*, we refer the reader to CC's index; for example, outside of B mathematics, are found D:5 for 'specification, quantity' under engineering, and a series of divisions for 'abnormal number' in G:4715, I:4715, K4715, and L4715, facet [E] with [2P] of biology, botany, zoology, and medicine, under pathology.

Measure, or measurement, is found generally in the energy facet, under frequently varying divisions (often 2 or 3, but also 6, 76, 78,8 and 1 for chronology, B9:1), but it becomes 'personality' for the measurement of distances in mathematical geography (U114); thermometry is in C4:2 and calorimetry in C4:3; electrical measurements are placed in C6:3 (in the energy facet of the canonical divisions relating to heat and electricity respectively, under physics); in engineering, the dynamometer is somewhat of a 'lost child' as 'personality' in D6296; under D66 industrial

electricity, is found under the energy facet a division 78, measurement (forming a part of 7 construction, execution)—but also, p. 2.46 of the 'divisions of Common Isolate digit e' schedules which apply to D66 industrial electricity, and enumerate measuring instruments, without, however, providing any further explanation on the special use of the 'e' (which is normally 'formula' see p. 2.4) than a brief paragraph p. 1.92 of the rules—where it is stated that a more complete schedule will be provided in volume 2.

Quantitative analysis is under E:34—it applies also to class F, so-called 'technology'—which relates rather to industrial chemistry. There are no divisions for measurement in the schedules for the biological classes (G to L); clockwork is found in MB9. Psychometry is in [2E] of S, under 6, and measurement of sensation under [2E]8. Under education, T:5 is 'educational measurement'. Prices are in X:76, but currency in X61 (personality). In the rather developed X:8 scheme for organization, the study of the time element is not mentioned (probably under X:89M1 motion study); accounting is in X:8F to X:8L, job rating in X:916, salary scales in X:931 to 934. Econometry is not mentioned, nor is sociometry under class Y, where there is a 'measurement' (Y:22) which seems to relate exclusively to anthropometrical measurements.

Chance, and the related notions of probability, luck, appear in R198 probability, under logic; under B28 statistics (a canonical division of algebra); under L:51 public hygiene statistics, and in MY32 games of chance; in the history of any one subject, a lower case *s* can be inserted after the geographical division, for denoting periodical statistics, or after the chronological division, the same *s* will then mean isolated (see p. 2.5 of the schedules). Statistical mechanics is CN2, in the 'systems' (i.e. the theories) of physics. Biometry does not exist, nor does statistics exist under economy (X) or sociology (Y).

In CC's index, *control* is mentioned in connexion only with the control of documents and the control rooms of electrical power stations; under 'regulation' is found a cross reference to L:4,642, for dietary prescriptions in medicine, and at L:576 regulation of habit in personal hygiene; 'regulative' gives a cross reference to the breaking of regulations. . . . Cybernetics is not mentioned. Finally, at 'governor' (Watt's old word), is found a cross reference to D6291 (under 'principle of mechanism and special device') entitled 'Governor. Regulation and control'—with no other division. The CC here seems to be singularly incomplete and behind the times.

It is feared that this examination of the CC may have proved to be

very fastidious. However, it may not have been unprofitable to contrast the theory of this classification and its practical realization. If the first seems to be very ambitious, with a strongly marked philosophical—and even dogmatic—tendency, the second appears, when one goes into detail, to be singularly empirical and often very arbitrary.

In relation to the UDC, a certain number of improvements are noted, probably greater flexibility, more 'hospitality', but it presents also some great gaps, especially in the classes relating to the physical and natural sciences, and to their applications. The 'classificatory language' of the CC—to use Ranganathan's own expression—is poorer and at times deficient (measurement, regulation). It allows much room for the individual initiative of the classifier³⁷—who can, of course, almost always 'manage' with empirical devices, such as the 'chronological device'; but this freedom runs the risk of leading to anarchy in the event the CC were used on an international scale and for the detailed classification of documents.

Certainly, '*la critique est aisée, mais l'art est difficile*'; it is always easy to find faults, inconsistencies, gaps, in the CC, as in all other classification systems. However, one must grant to its author certain exceptional merits: the CC is the first basically 'multidimensional' bibliographic classification which has ever been conceived. Ranganathan has played a pioneering role; without his work a great part of the movement of ideas which has manifested itself in recent years in the field of bibliographical classification would probably not have taken place. Particularly, he has found disciples in England who are perhaps not always faithful, but who have been inspired productively by his precepts. As chairman of the FID/CA Committee on the General Theory of Classification since 1951, he has given the committee considerable stimulus. Together with that of Otlet, his name must be placed in the front rank of those who are worthy of the respect and admiration of all classifiers.

14 After the Colon Classification

The CC seems, in fact, to have been the last new general classification system of international significance to have seen the light of day.

We must, however, set aside the question of Soviet classification systems, on which the information at our disposal is insufficient.³⁸

It could have been anticipated, about 1944-45, that Gérard Cordonnier was about to provide us with a new detailed general system. But he has never extended his universal classification

proposal beyond the first 25 divisions, published in his lecture to UFOD in December 1943 (see Note 2 in our Introduction) and which are reproduced (with a few minor differences) in his lecture of 11 April 1945, given at the Institut Technique du Bâtiment et des Travaux Publics.³⁹ He stressed also the need to provide what he termed 'functional endings', furnishing 'at least an elementary grammar . . . to distinguish names, adjectives, verbs, subjects and direct or indirect objects or complements' and added 'These functional indicatives will form the object of a special division of the classification, relating to particles of relationship'.⁴⁰ This would provide a 'universal terminological code'.⁴¹ These ideas—which approximate very closely those of more recent authors—have inspired Robert Pagès, as will be seen later. But this did not result in the development of any general classification system.

While a few organizations for specialized documentation adopted Cordonnier's method of symbolization—we shall examine below (p. 65-67) one of these special classification schemes, that of the Centre de Documentation des Constructions et Armes Navales—Cordonnier himself later provided a series of *categories*, of which he has never to our knowledge published a complete list, to be applied in conjunction with visual superimposed cards which, though not invented by him⁴² he had helped to popularize; the most detailed list of these categories is provided in the report which he presented at the Dorking Conference⁴³ and which we reproduce below:

Organisms and services (origin, or subject, of documents)
 (nature; places)
 Persons (miscellaneous categories . . .)
 Individuals (living beings . . .); biological conditions . . .
 Bodies (natural; simple, compound; . . .); miscellaneous conditions
 Miscellaneous equipment, (property, fittings, . . .)
 Miscellaneous actions, (physical; technical; economic; . . .)
 Intellectual concepts
 Documentary forms . . .
 Time . . .

In the report which he prepared for the Cleveland conference (earlier mentioned in our Introduction), Cordonnier reverts to his proposal for the creation of a synthetic terminology, by a polydimensional classification system, for which he proposes the neologism of 'syntermation'. But he adds (paragraph 0,25 of the summary): 'A complete project for meta-language cannot be proposed until after terminology has been placed in order in all of its specialty fields, logically analysing all distinct characteristics.'

Henri Clavier became the first president of the General

Classification Commission of the FID. In his three works published in 1942 and subsequently in 1956-57,⁴⁴ he pleaded the cause of a general system of documentation and proposed a classification scheme with a double notation (decimal figures and three-letter mnemonic symbols. There is nothing particularly new in the above scheme from our point of view.

Louis Glangeaud published in 1955 a very interesting article⁴⁵ outlining certain ideas on 'levels of organization'—of great importance in the establishment of a rational classification system—which meet those also of the Englishman Joseph Needham.⁴⁶ However, this relates merely to a basic, scientific outline—not a documentary classification.

In England, the Classification Research Group (CRG) is certainly tending towards the creation of a general classification. To date, it has produced specialized systems only, adopting Ranganathan's method of analysis by 'facets' (but not his fundamental categories). Nevertheless, the CRG seems to be heading toward the establishment of a general system, requested by D. J. Foskett⁴⁷ and by Barbara Kyle.⁴⁸ The latter believes that 'at this stage the principles on which such classifications [general] should be based are well established' and that only some financial support would now be necessary to establish a small secretariat or an advisory committee 'to draft a general classification' (p. 17). However, it does not appear to us to be so very definite that the basic principles of a general classification system are so 'well established'. In fact, the basic scheme provided by Miss Kyle in six parts:

- Natural phenomena
- Artifacts (i.e. products of human industry)
- Activities
- Special schedule of aims and purposes to be used in the social sciences
- Imaginative literature and music, perhaps also religious and philosophical speculations
- Form, space, time (common subdivisions)

on page 15 (with some detailed information on page 9) of her report, is a classification by 'categories', which, as a matter of fact, seems bound to create in practice a great number of artificial separations (an 'artifact' is often based upon a natural phenomenon and is used for an activity . . .). The scheme provided by Vickery (*CIS*, p. 20), on the other hand, relates to ' . . . "things" defining subject fields', from elementary particles to products (material and mental) of man, and its structure is derived to a large extent from the conception of 'levels of aggregation' the history of which he traces briefly in appendix A of his book (p. 127-9 and

p. 138-9). Subsequently, Vickery attempts to reconcile a classification by levels of organization with a determination of fundamental categories inspired by Ranganathan, although considerably more subdivided (p. 163-77), such levels of organization being related to the 'substance' category or the 'personality' category. (In our opinion, the success of such an undertaking is more than doubtful.) And Foskett, who seems to be much more critical as to the applicability of general categories,⁴⁹ seems to be moving towards the use of the concept of levels of integration.

2 General categories and the expression of relationships in specialized classification and coding systems

Simply for the sake of convenient presentation we shall group the various classification and coding systems examined in this chapter in geographical order. We could equally as well have arranged them by fields of study, but the resulting order would certainly have been more artificial. In fact up till now each country has not exchanged its 'classifying experiences' with others; as a result a small number of national 'schools' have emerged, which it would seem appropriate to examine in turn. We shall begin with the 'French school', not by reason of the accidental fact that France is our own country, but, because chronologically it would seem that its specialized classifiers have been among the first to seek new methods.

21 France

Here, as in the other countries, we shall examine only a relatively limited number of specialized systems which have seemed to us the most interesting from the standpoint of the study of general categories and the expression of relationships. This has led us, not only to dismiss all traditional systems involving a hierarchical classification or the usual type of alphabetical subject headings,

but also numerous 'non-conventional' systems which do not include a special study of the problems under consideration in the present report. This will explain, for instance, the absence of anything but a simple reference to a classification system such as that of the Association Française pour l'Etude des Eaux,¹ despite the interest presented by its schedule with a double entry: five columns for the origin and natural levels or tables of water, the supply of water, water and life, water and industry, used water and its rejection; four rows for generalities, legislation and regulations, scientific questions, technical questions and practical applications.

211 *Serge Tchakhotine*

The first author we should mention is a Russian citizen—subsequently a resident of France—Serge Tchakhotine, a student of Pavlov. In his book on scientific research is found a paragraph on 'scientific language and writing'² where he outlines a 'lography' method, the bases of which, he states, were dictated to him by the theory of conditional reflexes. It will perhaps be of interest to reproduce here the passage containing an explanation of 'his principle applied to Biology':

'The fundamental notions are divided into groups, e.g. the notions relating to morphology, chemistry, normal functional notions, pathological functional notions, etc. Each group has been assigned a characteristic symbol, an underscoring dash (—) for technical notions, an underscoring crescent (∪) for morphologic notions, a dot over letters (·) for chemistry, etc. The actual notions are represented by two letters, selected according to the mnemonic principle: e.g. the viscosity symbol is V_s , of the Cytoplasm Cy , of Cytolysis Cy , of proteins Pr , etc. Relationships are indicated by special symbols, e.g.: "action on" by a horizontal

line dividing the symbols: "action of acids on the core": $\frac{Ac}{N}$; the "function of" by parentheses: "viscosity of the cytoplasm": $V_s(Cy)$. Therefore, in order to define a problem, for instance the following (expressed in ordinary handwriting), "ultra-violet ray action on the viscosity of the cytoplasm of an egg", the following

would suffice: $\frac{uv \text{---} \leftarrow^{\circ}}{V_s(Cy)(Ov)}$.

It will be readily acknowledged that the ideas outlined above bear a close resemblance to those of Selye (see below, p. 101). Tchakhotine adds, moreover, that this lography, 'a system of written symbols which must correspond to the elements of present scientific thinking and to their functional relationships',

does not yet constitute 'a writing with symbols corresponding to the logical entities of a future Phenography', since he 'has been compelled temporarily, in order to test the method', to 'use current scientific notions'. He explains his meaning of 'phenography' by a criticism of the principle of artificial languages, such as Esperanto: 'this language, although simplified and rectified, is based upon the same principles as our modern languages, and is therefore vitiated at the very outset in so far as a faithful expression of thought is concerned. It is possible that the solution lies rather in the creation of a system of symbols, a special script—Phenography—the symbols of which would correspond exactly to the elementary phenomena or complexes of thought.' And he would like to see placed on the 'agenda a general revision of rules governing the formation of scientific concepts—true basic elements for the rational advance of research'. It is to be deplored that Tchakhotine, who thereby shows himself to be a precursor of much research at present in progress, has not, it seems, studied further along those lines.

212 Z. Dobrowolski

Z. Dobrowolski, a Polish engineer who was employed by the documentation service of the Office Central et de l'Institut de Soudure de Paris, 1942-43, at that time prepared for these organizations a classification, which is original in character by reason of its notation system. It uses very broadly and systematically the principle of the UDC's 'extension' symbol / to express groups of classes and completes it by a method of forming symbols which reserves for those most frequently used a group of two letters only; the symbols of three letters corresponding to classes of a higher hierarchical level, which include as subdivisions classes beginning with that indicated by the two first letters of the number and continuing with that indicated by a number made up of the first and third letters. For instance, WHT being 'work', the divisions of this group are those which go from WH, labour or man-power, to WT (the latter is temporarily vacant for future expansion; the symbols when completed end in WS work organization).

Dobrowolski's classification system is still in use and is used for the classification of the *Bulletin de Documentation de la Soudure*, published by the Institut International de la Soudure. It constitutes a 'faceted' classification system—by categories—which establishes groups of headings for fuels used in welding, the material, the processes, the applications (with two sub-categories: materials and manufacture), initial products, 'general problems' (metallurgical effects, restraints, conditions of operation, influence

of the medium), properties, study and control, industries and organizations. The relationship symbol : (borrowed from the UDC) is constant.³

213 *The Desaubliaux code for the Carrel Foundation*

Robert Desaubliaux prepared in 1942-43 a code for the documentation of the Carrel Foundation (abolished in 1944; the Institut National d'Etudes Démographiques subsequently took over part of the work done there).

To our knowledge, the code has never been published and exists in typewritten form only.⁴

It consisted of a code with a decimal notation, used in conjunction with pre-punched cards, centrally perforated according to a model established by the author (later commercialized by the Dequeker firm). The notions to be coded were grouped into eleven classes, the latter being distributed among three large divisions: physical medium, man, human communities. The most interesting aspect of this code, from the standpoint of the present study, was the existence of a certain number of series of common subdivisions, each applying to a specified field, under the name of 'modalities'. For instance, within the groups 01 to 03 (constitution of matter and structural universe), appeared modalities such as mathematical calculations, formulae, symbols; experimental method; measuring and testing equipment and instruments; industrial application; then four subdivisions relating to the various repercussions on man. The same modalities, with variables in their titles, appeared again in groups 04 to 07 living substance and 'organized media' (i.e. astronomy and geology). Under group 12 study of man from the standpoint of the human species, was a first series of modalities for physiology and a second series for pathology; parallel modalities (identical for those relating to age and sex) exist for groups 20 to 24 the conceptual universe of man. Under groups 250 to 254 genesis, followed by structure and functions of organized communities, the modalities relate particularly to the repercussions on and applications to various categories of social groups. Under the groups 40-46 professional communities (classification by branches of industry and activity), the modalities correspond roughly to analytical subdivisions of point of view of the UDC (by .00): raw materials, machinery and tools, operations, questions relating to the workers, organization, etc.

The Desaubliaux code corresponds, therefore, to a 'faceted' classification; this was merely in the nature of a coincidence and not an influence, for it had not been inspired by the ideas outlined by Ranganathan.

It will be observed that this code mainly avoided naming the

sciences, and was presented rather as a classification of 'things', activities, etc.

214 *Specialized classification systems inspired by Cordonnier*

A number of French specialized documentation organizations adopted, in 1943-45, the classification methods advocated by Cordonnier, especially from the point of view of notation (base 5, alphabetical notation, syllabified in groups of two consonants framing a vowel, with rules restricting the choice of combinations in order that the 'echelon'—i.e. the hierarchical rank—should remain visible in case of abbreviation of the number, use of the symbol ' which has the same meaning as UDC's ;, the adjunction of a symbol for the 'cyclical branches', i.e., the hierarchized divisions of a number formed by the combination of several simple numbers). Among the organizations adopting the system were: the Institut Français de Recherches Fruitières Outre-mer (IFAC), but this organization, finding it too rigid, abandoned the system in 1956 in favour of an alphabetical index by materials;⁵ the Institut Technique de Bâtiment et des Travaux Publics;⁶ and the Centre de Documentation et d'Instruction de la Direction Centrale des Constructions et Armes Navales (Ministère de la Défense Nationale).⁷

Didelin's classification system for the ITBTP is divided into 16 main classes, under 4 general groups; this order is linked to the notation system used:

- B general concept of works: programme, composition, aesthetics, archeology
- C necessary knowledge for the establishment of the technical construction project: resistance of materials, mechanics of fluids, geophysics, 'general conditions'
- D technique or art of construction: knowledge and general techniques, the actual construction, related installations, methods of execution.
- F work of construction: elements thereof, construction work directly linked to the life of man, linked to man's activities, 'external incidences' (comprising, for example, modifications, demolitions, disorders).

As interesting as his details may be from the standpoint of building techniques, this system does not seem to warrant retaining our attention further here.

The classification system of CAN-DOC (Constructions et Armes Navales) is divided into 23 main classes, distributed among 5 groups:

- B functional point of view: origin and destination of the documents, documentary format, aim or purpose of the document
- C social and humanitarian point of view: man, societies, legal standpoint, economic and financial standpoint, military art, and international relations
- D sciences, arts, letters: philosophy, art, and literature; pure and applied mathematics; sciences of matter and the forces of nature; sciences of the earth and universe; life sciences
- F technique, products, materials and elementary construction: industrial techniques (except machines and electricity); machinery; electricity, optics, acoustics, telecommunications; 'questions common to various techniques'
- G marine: general organization, services; ships; arms and war materials; aeronautics; earth installations.

Here again, the classification scheme is of greater interest to the specialists than from the general point of view concerning this report; it is, to all intents and purposes, an enumerative classification scheme of a rather traditional hierarchical type, with a syllabified alphabetical symbols instead of a decimal notation.

Certain headings are found therein presenting general categories: e.g., in DiB M relative values and variations of phenomena; FaB M relative positions; and in Fu 'questions common to various techniques', e.g. FuF faults or defects and inverse qualities. But, more often than not, the same notion is found in multiple locations of the classification under direct hierarchized divisions.

Let us take for example the case of notions relating to form. Geometry is in DeB; geometrical forms in FaB M, in a division entitled 'common points of view' of technique, FaB (which should not be mistaken for Fu, already mentioned above); structural properties in FaB Jo (under FaB J properties, a division of the foregoing FaB). Deformations are in DiC LaL (under DiC L mechanics of true solids); forming in FeC Je (Fe being 'techniques (except machines and electricity)', FeC manufacturing techniques, techniques of assembly, treatment, and protection and FeC J manufacture of metal and plastic objects); the tooling or working of form in FeC JiL.

Another example relates to nuclear questions. Nuclear reactions, fission, are found in DiG L corpuscular physics; the machines employed in nuclear physics are in DiG Lo. But 'nuclear machines' are found at FiG L, under a Fi division 'thermal machines, mechanical, hydraulic, and pneumatic machines', and nuclear explosives in GiB N.

Regulatory machines or instruments are found in FiB LiM, but the theory of regulation is in FuC Jo.

Under FuC L is found a heading 'use, application, exploitation or operation', but the applications of photography are in

FoC PiJ, those of cinematography in FoC PiL, and those of electricity in FoB La.

Tests have been assigned a heading in FuD, which seems general; however tests of ships have a special number GeD L.

Optics is under FoC, in a most composite group, F technique, which bears the title 'electricity, magnetism, radio, optics, acoustics, telecommunication'. But optical dimensions or sizes are in DiB Lo, and the constants and optical properties are under FaB JaP (in a FaB J which groups—in principle—all 'properties').

Lighting is in FoC N, a division of optics; but electric lighting is in FoB LuJ, a direct division of electrotechnique: as seen here, it seems hardly worth while to have a symbolism for the purpose of expressing relationships, if it is not used.

Meteorology is in FeF M, under FeF optics and high precision technique—where, moreover, optics is not classed, but mentioned there as a reminder only; however, measuring instruments are found in FuD Le, and the recording of measurements under FuD Lo.

In so far as notions relating to temperature are concerned, the dispersion is even greater. A list, probably incomplete, is given below:

DiB Le	thermal sizes
DiD J	thermometry, calorimetry
FaB JaM	thermal constants and thermal properties
FaB JuL	relative temperature
FeC PaJ T	fireproofing, protection against heat, insulation
FeC PeJ R	thermal treatments and thermochemistry
FeD LaL	action on temperature (in 'technique of habitability')
FiC	thermal machines
FiD	thermal exchangers
FoB LuL	electric heating
GeC LeP	air-conditioning installations (ventilation, heating, etc.) in ships
GiB Je	freezing temperatures, fusion temperatures, etc., of powders and explosives
GiB JeM Ra	potential, explosive heat, formation heat, of explosives
GiB JeM SaR	temperature coefficient of explosives
GiD La	dispersion caused by heating of products dispersed
GiF Je	thermal and incendiary action of weapons
GiF L	protective clothing against heat

Moreover, it should be noted that as a general rule there are no cross references from one heading to another and that the classification scheme has no alphabetical index, therefore it can be deduced that indexing must be a very laborious proposition involving risks of error.

It is difficult to see what advantage a 'multiple coordinate' classification of this type can offer over a system of alphabetical headings of subjects with a good cross reference system, or even over the UDC, and it is probably not a particularly faithful application of Cordonnier's theory, which anticipated that the 'complex notions' would not be 'introduced directly' into his classification system, 'since they are derived from the combination of a number of points of view' (conference at the ITBTP, Circular X, 5 *bis*, p. 12).

More recently, Pierre Rennes, chief of the Studies Department of the Centre de Psychologie Appliquée (André Vidal et Cie) has elaborated for the documentation of this company a very ingenious coding system, to be used in conjunction with visual superimposed cards of the Sélecto-type (Cordonnier); if he received inspiration from Cordonnier's idea, he did not, however, use either one of his symbolization systems.⁸

The CPA (Centre de Psychologie Appliquée) code has not been published; we mentioned it briefly in two reports in 1958.⁹ We quote below the paragraph devoted to the code's 'general principle' in a letter which we received from Mr. Rennes dated 11 April 1958:

Documents to be classified are considered as points located in a space of n dimensions, the dimensions being represented by categories of concepts of closely related nature. Classification will be so much the easier for those categories which are homogeneous and independent of each other.

Practically speaking, the following procedure was followed:

- (a) We have collected among a certain number of general publications relating to our specialities, words of current usage, chosen in the alphabetical indexes of these works.
- (b) These words were regrouped into concepts, with this particularity that the terms indicating opposites find themselves regrouped under the same concept (present - absent; masculinity - femininity; etc.).
- (c) These concepts were regrouped into a certain number of hierarchies of content as homogeneous as possible and corresponding to the point of view by which one may be called upon to retrieve documents. These hierarchies constitute the 'dimensions' of our documentary space.

There follows a list of the latter:

Knowledge - sciences	Relationships
Fields of application	Processes
Functions	Operations
Methods	Representations
Instruments	(graphic and mental)
Behaviour	Documents
Anomalies	Agents - media
States	Places - organizations

Countries
Groups
Individuals
Individual characteristics
Individual traits
Anatomy
Qualities
Dimensions and sizes
Indicia
Concepts

Under a fourth point, P. Rennes goes on to explain the assignment of numbers to concepts, of which approximately 120 have received 'monovalent' numbers (each identifying one concept only) and approximately 800 have been assigned 'haphazardly, as far as possible' over 200 'polyvalent' numbers.

In order to provide a few examples, a category such as 'Processes' (CS) is divided into nine subgroups (evolution; organic process; regulation; perception; sensorial processes, gnosis; action of motory nerves; overheating; fatigue; memory, retention) to each of which has been assigned a number of two letters or two figures, or one letter plus one figure: evolution, for instance, becomes GS-98, and perception CS-PP. Most of the subgroups are in turn divided, and eventually subdivided: for instance evolution, CS-98, gives adaptation CS-98-AD, which includes CS-98-AD-29, 'acculturation', and CS-98-AD-SL, social adaptation.

Under the sensorial processes, CS-PP, hearing is CS-PP-OD, and vision CS-PP-VS. There are never more than four hierarchical rows.

Combinations of letters are used, for example, under instruments UT, to audio-visual instruments was assigned the combination of letters UT-OD-VS (OD being audition in the 'processes' and VS vision, as we have just seen above); measuring instruments being UT-MZ, audiometer will be UT-MZ-OD.

It is hoped that this interesting system will be published in the near future.¹⁰

215 *The Robert Pagès coded analysis*

Since 1948, Robert Pagès, at present chief of the Laboratory of Social Psychology at the Sorbonne, has elaborated a method referred to by him as a 'coded analysis' which owes much to Cordonnier, but which later developed independently. It has been partially outlined by its author in various publications¹¹ and we have on numerous occasions underlined its interest.¹²

It is not possible to give here more than a very brief general idea of these coding methods, which the recently published article in *Chiffres* outlines in some detail; any interested reader will find it easy to refer to the original publications.

The Pagès code offers itself as an alphabetic-numerical code in its present form; it had previously gone through a phase where symbols and accents, and upper- or lower-case letters were used.

The numbers are set aside for the expression of a certain number of 'syntactical' relationships (other relationships, as will be shown, are expressed by 'morphological' methods), or to 'introduce' the parts of the code word which are not constructed with the aid of the 'lexicon'—i.e., of the systematical code of concepts; finally, two numbers have a particular use.

The syntactical relationships stand, primarily, for union and intersection, taken more or less in the same sense as the elementary theory of aggregates (reunion and intersection); they are symbolized respectively by -5, -4, or -3 and 4, 8 or ∞ —according to 'levels' and 'calibres'.¹³

Two other relationships are expressed by syntax: 'thematization' and 'subordination'; they are asymmetrical and each has two aspects.

The two aspects of thematization are the following:

thematized = subject to, treated by

thematizing = bearing upon

According to the levels, they are assigned, respectively, symbols -2, -1 and 5, 9.

For example, *ip* representing in the lexicon an opinion or evaluation, and *aj* theory, one obtains: $ip^{-2}aj$ = theory of attitudes and opinions; *i* being psychology and *ame* concept, one obtains ame^5i = conceptualization of psychology.

The two aspects of subordination are the following:

subordinating = utilizing, making use of

auxiliary, subordinated = used for

According to the levels they are indicated respectively by symbols 6, 91 and 3, 7.

For example, *ane* representing ordinal indication, qualitative estimate, and *ya* order relationship, one has ane^6ya = ordinal scale (i.e. a qualitative estimate using the order relationship). Inversely, one would write ya^3ane = order relationship used for a qualitative estimate.

A series of ten numbers serves to indicate the 'segments' of symbols relating to proper names or to dates. The latter are preserved, in general, as they appear, the abbreviation 14' = XVth century being acceptable for the centuries; proper names relate to persons, organizations, languages, implantations, or

places; they are taken in their language of origin and abbreviated according to the rules outlined in the article in *Chiffres*, p. 109-11. Persons studied are indicated by 05, authors by 0, and organizations by 06; organizations as sources of documents by 01; languages studied by 07 and languages of expression by 02; locations under study by 08 and place of origin of a document by 03; finally, periods studied are assigned number 09, and dates of issue of documents 04.

As we have indicated earlier, two numbers play a special role, i.e., first, number 1 which indicates a part of a number relating to documentary form. Documentary forms, as a matter of fact, are derived from the lexicon. For instance, *jüic*, in the lexicon, represents a 'temporary institutional group; assembly, meeting, congress, etc.'; ¹*jüic* will represent: document produced by a congress.

The second number serving a special purpose is number 2, or 'analytical prefix': omitted at the beginning of a symbol, it indicates, within a complex identification, the termination of a segment of a proper name, of a date, or of a documentary form, and passage to a segment derived from the lexicon and relating to the subject of the document.

The *lexicon* consist of 22 categories of notions:

- a* science, scientific knowledge, cognitive activity or knowledge acquired
- b* multiplicity aspect, aspect of organization or of characterization of a multiplicity
- c* individual (= singular, not plural), element, more particularly the human individual
- d* beings and logico-mathematical properties
- e* facts and physico-chemical properties
- f* field of concrete natural sciences, with the exception of those relating to human beings: astronomical facts, earth, minerals
- g* living beings, life
- i* conduct or process which is translated by a behaviour effect in a living organism
- j* group, social phenomena (including extra-human)
- k* pragmatic aspects in general, complex action
- l* communication: conduct, element or fact of communication
- m* technical operations principally for physico-chemical application
- n* technical operations principally for living application
- o* social practice: economic phenomena, power, use of armed force, ethics and law, education
- p* pleasures, games and arts, hedonic, ludic and aesthetic social activities
- r* metaphysics, religion, myths, magic
- s* space (abstract or concrete, real or metaphorical)

- t* time
- v* regulation, factors governing change
- w* property of things, character
- y* relationships and 'relations' in general, excepting statistical associations and logico-mathematical relationships (see under *d*)
- z* negation of

It will be observed that we have indicated here by a separation between *r* and *s* the existence of rather distinct groups of these categories. In effect, those from *s* to *z* group together 'suffixes' which, without prejudicing their role of ordinary 'morphemes' (i.e., isolated words), 'serve to indicate the particular *predicates* to one or more variables'. Excepting any rule to the contrary (which is found, for instance, in the *g* and *o* categories), letters from *s* to *z* are not used in non-initial position, and morphemes of which the initial is *s*, *t*, *v*, *w*, *y*, or *z*, can, therefore, as a general rule, be combined directly with the others. For example, *i* being behaviour, conduct, *iw* will indicate properties of behaviour and *yeb* being a determinant, *iyeb* will be the determinant of conduct.

We really have here, this time, a synthetic system of coding: Pagès' lexicon contains a minimum of traditional hierarchical series; these are encountered mainly, as he states, in the 'far regions' of special points of view of social psychology 'where it would have been both hazardous and useless' for him 'to go beyond the elementary and traditional'—especially in category *g*, but also in *m* and (which is somewhat surprising) in *o*. Generally, the morphemes of the lexicon correspond to elementary notions and result, as indicated by the author, from a 'study of the coded definition of thousands of notions in humanities'.¹⁴

Although the Pagès code was not established for general use, it could be usefully consulted for a large number of general categories of ideas: for example, for those which he placed under *a* to *c*, *k*, and *s* to *z*. Since we cannot reproduce here all of these series, we will limit ourselves to those of *y*, which correspond more nearly to the object of the present study, according to a text kindly made available to us by Pagès:

- ya* Field of *so-called order relationships* (notions incorporating a qualitative or quantitative point of view are formed by a combination of *ya* with *wa* (qualitative property) or *we* (quantitative property), as the case may be. For example: *yab* signifies *ex-aequo*, *yabwa* will be homogeneity, qualitative parity, and *yabwe* will be equality, quantitative parity). Indicates, in addition, comparison in general
- yab* *Ex-aequo*; equal to

- yac* Before, in a relationship order; e.g.: 0 is prior to 1 in the series of integers; smaller than. (Symbol with a general meaning, not necessarily quantitative.) The special, chronological, dimensional meanings, etc., can be specified by combination
- yad* Intermediary between
- yaf* First among . . .
- yag* Last among . . .
- ye* *Determinative relationships*
- yeb* Determinant of, with causative effect on (determinants which are not specified as principal or secondary, indirect or direct, diffuse or well defined in their action). A cause
- yebe* Determinant of (outside the system considered)
- yebi* Determinant of (inside the system considered)
- yec* Indirect determinant, influencing, conditioning in a secondary or diffuse manner; influence in that particular sense (and not in the sense of the ascendancy of a person)
- yed* Giving birth to; producing. Source in time; origin; initiative; starting; set in motion. To be distinguished from *tam* (Commencement, beginning of a process, without consideration of external active intervention, as opposed to *yed*)
- yef* Active interruption; which puts an end to; suppression, partial or total destruction (injury)
- yeg* Serving as a framework of reference; serving as a centre for; serving as reference mark or system of reference marks; which serves as a foundation for (in that sense). 'Founded on' will be identified as *yegü*
- yi* *In concrete relations with. Concrete relationships*
- yib* Accompanying; simultaneously present (an idea frequently expressed by the prefix 'co-'). Concomitance
- yic* Serving for; principal means of; process of realization, of locomotion in Lewinian 'topology'; 'facilitation'. This notion of mediation toward an objective is expressed also by *yif* which designates means of contributing. The *yic* notion as well as the *yif* notion can also be expressed by the converse of the 'auxiliary relationship' which does not distinguish between principal means and contributing means (cf. grammar in *Chiffres*, p. 117, and above, p. 70 'subordination')
- yid* Forming an obstacle to; barrier to; preventing (without conveying the idea of combat; relates only to an orientation of conduct which is blocked by the obstacle) Obstacle; barrier; obstruction
- yif* Reinforcing; aiding (idea often expressed by the prefix 'pro-'). Aids (personal or social); reinforcement of . . .
Auxiliary
- yig* Supplying, transferring. Transfer of; supply of
- yiga* Supplying without compensation. Supply without compensation (e.g. gift)
- yige* Supplying with compensation. Supply with compensation (e.g. sale)

- yij* In competition with; in rivalry; indirect struggle in the pursuit of common objectives
- yik* Forcing; constraining. Active restraint exercised upon . . .
- yil* Fighting against, corresponds fairly closely to the sense of the prefix 'anti-'. Fight
- yim* Attacking. Attack; aggression; initiative of combat
- yin* Resisting; defending against. Resistance (as opposed to an attack); defence (as opposed to initiative of combat)
- yo* Capable of (that which precedes). Capability of action; potentiality; virtuality in that sense
- yob* Capable to a high degree. High degree of capability
- yoc* Capable to an average degree. Average degree of capability
- yod* Capable to a low degree. Low degree of capability
- yu* *Mutual relationships* (cf. note, paragraph 3, below).
- Reciprocity

yü *Converse relationships* (which are derived by infixation)

Note. *Syntactical reminder:*

1. The *y* has two meanings: a meaning 'relative' to the suffixed radical, e.g., *xyeb* = determinant of *x*; an 'absolute' meaning separated from the other [in the above enumeration] by a hyphen, e.g., *yeb* = a cause
2. Converse relationship: *yü*. The converse relationship of *y* : *yü* is such that *h y q* is equivalent to *q yü h* (cf. grammar)
3. Reciprocal relationships - Reciprocity, 'relationships between' are indicated by the addition of the letter *u* to the number. For example, *yeb* signifies determinant of; *yebu* signifies determining among themselves. This symbol, and that one only, changes to plural the preceding symbol. For example, *xyu* signifies relationships between the various *x*.

The above schedule calls for certain observations.

In *se* and *si* there are 'external' and 'internal' notions; it would probably have conformed better with the general character of the coded analysis (the decomposition of complex notions into simple notions) to express 'external determinant of the considered system' by *yebse* and 'internal determinant of the considered system' by *yebsi*. Likewise, *wec* and *wed*, in *we* quantitative property, respectively mean 'to a high degree' and 'to a low degree'; it seems that it would have been preferable to form the notions here coded as *yob* and *yod* by a combination with *wec* and *wed* rather than by direct division of *yo*; but it would have then been necessary to provide in *we* for a coding of 'to an average degree'.

We must admit that we do not clearly see the difference between *yic*—a morpheme of relationship—and the syntactical relationship of auxiliarity, 3 or 7; despite Pagès' explanation, according to which the latter 'does not differentiate between

principal means and contributive means', it is not clear to us why in one case a process is used which falls under the head of syntax, and in the other case a morphological process is used. It is true that this would raise again, more generally, the entire distinction between morphology and syntax.

There is, in *bii*, a morpheme 'role, function . . . partial contribution in the process which leads to a total result; "co-operation in this sense"'; it seems rather curious that *yif* is separated from this *bii*, without even a cross reference from one to the other in the lexicon. Similarly, the *yad* to *yaf* morphemes seem to fall under the idea of classification; however, the notion of class, in general, is found in *bi* (and that of structure in *bu*). More generally, the separation of categories *b*, *c*, and *d* on the one hand, and *s* to *z*, on the other hand, could be questioned; it would, perhaps, have been preferable to group all these abstract notions at one place only, either at the beginning, or at the end of the code. This would also have obviated the separation of the *s* space, and of *di* topology and geometry.

216 *Dr. Jacques Samain's coding system*

For a number of years, Dr. Jacques Samain has experimented at the Centre de Documentation of the Centre National de la Recherche Scientifique (CNRS) with coding systems in the field of pharmacology, in conjunction with the use of his selection apparatus, the 'Filmorex', with a view to permitting the use, for retrospective research, of abstracts published by the *Bulletin signalétique* of the CNRS. No part of his code has as yet been published; we know only of a 'general outline' indicating notably one symbol (07.3 . . .) for 'interaction' relationship (specified, subsequently by a fourth figure for 'of': 0 microbes, with 6 physical agents, and 9 miscellaneous, and a fifth figure for 'on', with the same divisions: for instance, the symbol 07.360 will indicate interaction of physical agents on microbes). A series of 'point of view' divisions is also found: e.g., 07.03 -metry, dimensions; 07.09 applications; 07.14 physical constants; 0.721 physiology; 07.50 pathology; 01.600 study conditions (with divisions for theoretical study, .601, experimental .602, technical .608).

217 *The French Atomic Energy Commission's classification, and research on coding systems*

During the first years of existence of the Service de Documentation of the Centre d'Études Nucléaires de Saclay (French Atomic Energy Commission, CEA), the UDC system was used. Subsequently, P. Braffort elaborated, in order to replace this, an

alphabetico-numerical classification system, the first edition of which was published in 1953 and revised in 1956.¹⁵

The above system has a 'bidimensional' basis; the introduction outlines a justification for this choice, frankly a rather curious one:¹⁶ 'The filing equipment which is available to us imposes a bidimensional basis for our system; it consists of the rectangular shape of lines and columns formed by the card file drawers. In order to define these columns, we shall select by order of magnitude, i.e., the level of the phenomena under study, with a few modifications which will allow these columns to correspond more or less to the large CEA subdivisions.'

These five columns are therefore:

Inter-scale phenomena, communications, physics of the field
Corpuscular scale
Nuclear scale
Atomic and molecular scale
Macroscopical scale

This enumeration bears some resemblance to Cordonnier's 'cosmological standpoints' in his scheme of universal classification.¹⁷

'The lines', the report goes on to state, 'correspond to the different methods of approach to scientific and technical problems':

Theoretical problems
Production of phenomena
Measurement of phenomena
Description of phenomena
Utilization of phenomena

For the notation, 'to the square of 25 elements defined by these 5 lines and 5 columns correspond the square of symbols':

A B C D E
F G H J K
L M N O P
Q R S T U
V W X Y Z

But, as the author readily admits, this 'two-dimensional diagram, more flexible than a linear scheme, does not abolish . . . difficulties. The structure of matter, the human techniques for its discovery and utilization create a much more complex scheme than a double-entry schedule. There are problems which bring out phenomena at different levels (for example, the electromagnetic properties of

solids, chemistry under radiation, etc.), on the other hand, it is frequently impossible to separate the theoretical and experimental aspects. Such as, for instance, in articles on the resistance of materials, background noise in semi-conductors, etc.'

Braffort then divides the 'spaces' in the diagram, represented by a letter, 'by forming headings denoted by the coupling of two letters. Each letter corresponding to an element in a two-dimensional aggregate, a pair of letters correspond to an element in a four-dimensional aggregate. Therefore, since X contains the applications of nuclear physics ('radioelements being essentially in question'), the pairs formed with X 'should correspond to this definition'. Z, for its part, relates to industries, and XZ will be 'industrial applications of radioisotopes'.

This ingenious system however—as could be anticipated—corresponds with difficulty to the needs of practical classification, even with classical subdivisions included subsequently in the 'quadrimensionals' formed by 'pairs of two letters'; it was therefore subsequently necessary to compile a certain number of special rules which will be found on p. 6-7 of report no. 568 (repeated from p. 7-9 of the 1st edition). Among other requirements, it 'was necessary to introduce headings relating to the biological sciences, geology, and astronomy. Their place in our system constitutes a departure from the hierarchy of the structure of matter, the basis of our division by columns. Such departures are unavoidable unless a new multiplication of symbols and of their combinations is agreed to, which would not have been justified by the present importance of these disciplines within the CEA.'

We have quoted this Introduction at rather great length, because it shows clearly both the rather remarkable intentions of the classification scheme (notably in so far as a recognition of the 'multidimensional' character of concepts is concerned, with regard to classification by 'scale'—it is recognized here that the latter forms a part of a rather general, or even 'modern', current of thought), and its weakness in practice which is doubtless due above all to a much too rigid symbolism, whose artificial rules impose upon it a framework in which reality can no longer find a place except at the price of numerous 'twists'. In fact, the latter have multiplied in the second edition where, writes J. Iung, 'the special rules set down in the first edition were not all preserved'.¹⁸ This amounts to saying, practically, that—numerous disciplines or techniques still being considered of no great importance for the CEA in 1953 and not at all in 1958—new divisions have been 'inserted' wherever possible, without concern for the initial quadridimensional theory. Of this there remain of course some traces—the most obvious one being perhaps that, in the order of

the classification schedules, B comes after V, C after R, etc.¹⁹—but so obscured in practice that it is no more than a source of confusion.

For instance, one should not theoretically find under L, M, N, O, P 'measurement' only—and all the measures should have in their symbols at least one of these letters; however, spectrometers α , β and γ are under GR, spectrometers of mass in GS, the spectrometry of neutrons in HH60, the measurement of neutronic flow in HH50, the measurement of thermal conductivity in EK20. On the other hand, in LQ has been placed everything relating to the electronic microscope in general (LQ50); in MZ, one not only finds Geiger counters and scintillators used in mining prospecting (in addition, counters and scintillators in general are in MM, MQ, and MU), but also transport machinery and equipment for prospecting, method of prospecting, etc. (prospection is also found in YP).

In MJ, is found all that relates to photographic emulsions, photography and cinematography in general being under OO, together with a very large part of optics in general, under a heading entitled 'optical analysis', a part of O chemical analysis. Under OQ, one finds 'inserted' electrochemistry in general, and not merely potentiometry, amperometry, coulometry, which are in OQ60 and OQ80. In PO10 and PO20, under P macroscopical measurements, one finds X-ray equipment—including the production of the latter, PO11, although the 'production of phenomena' should be indicated by at least one of the letters F G H J K. Atomic piles in general are in HN, including the materials, protection, etc., all matters obviously far removed from 'measurement' (which are in HN20 experiments, measurement).

The 1953 classification system of the CEA has thus led to a complicated filing system, in which the 'non-initiate' is more or less lost—and in which it can be presumed that the 'initiates' themselves must encounter difficulties in finding their bearings.

In 1958 we conducted an inquiry in four research centres as to the habits and opinions of researchers. Of 174 persons who participated in the inquiry at the CEA, 142 replied to a question (14g) regarding the usefulness of the catalogue: for 46 this search was of 'great utility', and for 50 'of a certain utility', and for 61 'of little use' (on the other hand in 83 answers the documentation service was indicated as 'of great utility' for obtaining documents, and 'of little use' in that respect in 22 answers only). In response to a question (27) concerning the difficulties encountered by researchers in using the documentation service, 25 indicated difficulties in using the subject catalogue: 'complex', 'too crowded', 'unconsultable, too muddled', 'complexity and lack of logic', 'not very usable for the uninitiated', 'irrationally filed', 'crowded and

tricky to use', 'great loss of time in searching in the catalogue', 'location of the drawers' (this refers, no doubt to the particular alphabetical order we indicated above) 'too complicated to use', 'complex catalogue, usable only after acquiring the habit of searching in it', 'not sufficiently subdivided, at times arbitrary filing', 'insufficient classification' (this same answer adds, 'it is practically impossible to compile a thorough bibliography on any specified subject with the help of the Doc. catalogue', 'I could not make head or tail of the catalogue', 'unusable catalogue' (answer of a biochemist), 'better to classify the catalogue' (underscored three times), 'poor coding of articles in relation to a satisfactory principle', 'any material for which one has imperfect references is practically impossible to find'. A few other answers indicate simply difficulties encountered, without specifying their nature; one only is more detailed (it was received from an engineer in chemical engineering, a graduate of the Massachusetts Institute of Technology (USA):

Ambiguity often occurs in classification (e.g. descriptive chemistry and separation of isotopes); articles could sometimes be classed in one group, and sometimes in the other . . . I believe that the 'Fichier matière' [Subject Catalogue] could be simplified by eliminating the groups and subgroups [i.e. the 'quadridimensional pairs'] and by adopting a very simple classification system. For instance, at the Massachusetts Institute of Technology, the classification of the subject catalogue is alphabetical, with duplicate cards in various locations, if necessary.

Actually, it would seem that a classical system of alphabetical subject headings, such as that used by the USAEC,²⁰ despite all the criticism which may be levelled against such a method, would have presented certain advantages over the alphabetic-numerical classification of the CEA.

For that matter the latter is now turning in a totally different direction—that of automatic retrieval. Preliminary studies conducted by P. Braffort and A. Leroy, have been summarized in Note 278 of the CEA and the article in the *Bulletin des Bibliothèques de France* already quoted;²¹ one more is expected.²²

These studies have been based primarily on the use of characteristic *diagrams* 'constituted essentially by words linked by arrows bearing a specific syntactical meaning'. Leroy and Braffort single out four *categories of words*: objects (or entities), inserted in rectangles; properties and conditions, within circles; actions, represented by dotted arrows.

Furthermore, they use six *relationships*:

Action relationship, with no particular symbol other than the dotted arrow; the name of the action is inserted in full

Relationship: 'relating to' (object concerning another object, property relating to one object, condition qualifying an action).

Symbol: A

Relationship: 'results in' (result of an action upon an object, a property, or an action). Symbol: R

Relationship: 'by means of' (action performed with the aid of an object, of a property, or of another action). Symbol: M

Relationship of 'location' (place where an action takes place). Symbol: L

Relationship: 'for the purpose of' (action performed for the purpose of obtaining another). Symbol: B

Relationships other than those of action are expressed by solid arrows accompanied by the appropriate letter, and oriented as follows:

Property toward object, or object toward more general object, or condition toward action, for A

Toward what is produced, for R

Toward that which serves as a means, for M

Toward the location where the action takes place, for L

Toward the objective, for B

Each relationship can, moreover, be affected by a *negation symbol* (N) to indicate the absence of an object, of a property, of a condition, or of an action.

The bibliography contained in the 'Note CEA' no. 278 does not mention the works of Pagès; however, it can be recognized that relationships R and M are the equivalent of the 'subordination' relationship of Pagès and of the *yic* and *yif* of his lexicon; his lexicon, on the other hand, contains a morpheme *ve* which indicates the 'objectives effectively sought by an agent'.

It will be noted that relationships R and M are not independent: in fact, in the example mentioned by Leroy and Braffort (p. 12) for the R relationship, manufacture by UO_2 patch sintering, an R arrow is used from sintering to patch, although the example for M relationship (manufacture by sintering) shows an M arrow from manufacture to sintering. This would seem to create an eventual source of confusion in the application of the code.

In another passage of their study, the authors indicate that, in order to apply their method to an aggregate of concrete documents (on the technology of reactors), by using an IBM 650 computer, they have used 99 relations (p. 19)—of which they do not provide a list. Only a 'control' relationship and one on 'absorption' are found in the subsequent table (p. 22). It is therefore probable that the 'language of reactors' (as they refer to it) includes at present, for them, 94 'verbs'.

We shall not dwell here upon an aspect of such coding which

provides, for the conversion of diagrams into 'key-phrases', the addition to the words and symbols used of figures indicating the 'starting point of the arrows expressing the relationships', and serving the purpose as well, in a way, of 'punctuation marks' to separate the groups of terms (see p. 16-18 of the note).

One paragraph in CEA note 278 (p. 14-16) is devoted to the 'display of fundamental concepts'. This represents an operation which is identical with that which Perry-Kent named an analysis into 'semantic factors' (see below). The note remarks that 'a large number of the properties can be simply expressed, starting from actions', therefore 'the idea of solubility is directly linked to that of dissolution', and adds that 'the same applies to a great number of objects. For example, refrigerant: definable directly from the action of cooling'. 'This reduction' add Leroy and Braffort, 'can be pursued on the scale of the actions themselves', and he furnishes the example of 'cooling', an action which can be written 'cold' (adjective) with a dotted arrow surmounted by the symbol $\dot{-}$. Although this is not dwelled upon in the note, these facts may give rise to some doubt as to the validity of the classification of words in 'categories': objects, properties, conditions, actions.²³

Leroy and Braffort, in examining the limits in the reduction of the number of key words, write 'that a reduction process can easily be supposed which would result in replacing words by very ramified diagrams in which the purely nominal part would be reduced to a minimum: numbers, elementary particles', this reduction being compensated by the growing complexity of the diagrams, and they add that 'a reasonable compromise should be found between the number of key words used and the complexity of the diagrams. It would seem—without considering this as more than an indication—that about ten key words could suffice to express a particular notion, even a most complex one, without the associated diagram becoming overdeveloped' (p. 16). Here again they rediscover a problem recognized since Leibniz.²⁴

On the other hand, analysis by a diagram does not seem to be able to go beyond a certain complexity. The diagram by Leroy and Braffort for the article by Gordon, Stone, and Epstein, which they give on p. 21, facing the original 'synopsis', reveals in relation to the latter, a definite 'loss of information': notably no mention is made that the authors have attempted two distinct systems for the absorption of neutrons, one with a *tank of mild steel* filled with liquid cadmium, which is not practical because a large part of the reactor's power must be used to *pump* the refrigerant between the plates of the tank; the other with a *stainless steel* tank containing helium, connected to a *manometer* or pressure-gauge containing the liquid cadmium in such manner

that only a small portion of the reactor's power is used for the pumping. (We have underlined all of the terms which have disappeared in the diagram.) Leroy and Braffort indicate, as a matter of fact, that 'certain other elements could be added to the diagram in order to obtain a more complete analysis'. But, they go on to say, 'we then approach problems which are linked to the necessary degree of analysis, problems which can only be solved after numerous experiments'.

Braffort and Leroy have continued their studies since 1959-60 at Euratom, where they run the 'Grisa' group (Group for Research on Automatic Scientific Information), which has since become a section of Cetus (Centre for the Treatment of Scientific Information). 'Grisa' started with the ideas of Braffort and Leroy, but has developed also under the influence of certain American research (Harris, Chomsky, Yngve, and the syntactical theory of L. Tesnière) a method which should lead to the automatic analysis of scientific texts, called the 'programme of conflicts'.²⁵ In the framework of this method J. Ruvinschii has worked out a classification of words in three categories: action, properties and objects—replacing that of Leroy and Braffort (see p. 79, above) and has tried to draw up a list of relationships based on the analysis of French propositions, a list which consists of 18 groups.²⁶

J. Jung, at CEA, has also undertaken other research relating to the study of a system of automatic documentation in the field of controlled thermonuclear reactions; however, up till now he has published only a very general account of this study.²⁷

218 *Research by Jean-Claude Gardin*

Since October 1955, Jean-Claude Gardin, first at the Institut Français d'Archéologie at Beyrouth, and later at the Centre d'Analyse Documentaire pour l'Archéologie of the CNRS (Louvre), has been engaged in a series of extremely remarkable undertakings in the field of conceptual or formal analysis of documents (in the broadest sense), which unfortunately, to date, have only been partially published.²⁸

Gardin started from the established fact that the terms of natural language currently used for the detailed description of objects (weapons, tools, pottery, monuments decorated with figures, etc.) were too vague and not sufficiently standardized to be usable for the application of methods of mechanical retrieval to archaeological documentation. 'Compact codes' could be substituted thereto, analysing these objects with the aid of a number of notions with a very restricted basis, constituting 'strict systems, relatively simple for defining, within a certain field, the

specificity of a particular object and its formal relationships with allied objects'.²⁹

An 'analytical code' for bronze age *metal tools* from the Balkans to the Indus, was the first to be created (see the Beyrouth publication, 1956). It consists of a 'series of tables or schedules in which the various characters required for an exhaustive description of tools are grouped by heading': for the shape of the functional part, the kind of handle, the dimensions; the section and contour of the body of the tool, the connexion of the latter with the cutting part, the shape of the latter; the handles, casings, bristles, and heel; the ribs and ridges, various peculiarities, the decoration. Under each heading, a certain number of types are indicated; the aggregate of characteristics of the same tool are coded and transferred to the elements of appropriate equipment used in mechanical retrieval (at first, centrally pre-punched Desaubliaux Dequeker cards were used; Gardin at present uses Selecto cards).

Vases are the object of an analytical morphological description of their various parts, their dimensions, the type of connexion between the various parts, their respective proportions.

A code was later established for *ornamentation*; it distinguishes, on the one hand, 20 'elementary symbols' (dot, line, spiral, Z, loop, etc.) distributed in two parallel series, rectilinear and curvilinear, of 10 symbols each, and, on the other hand, approximately 30 'operations' and geometrical arrangements of elementary symbols (symmetry, rotation, etc.) distributed among six classes (polygonal, symmetrical, radial, linear, intercrossed, and intersectional arrangements). By thus combining elementary symbol with an operation a description of 'primary ornament' is obtained; successive operations engender 'secondary' and 'ternary' categories. Beyond that point, the eye no longer distinguishes individual 'ornaments', but 'ornamental compositions'; this transition is similar to that from the word to the phrase in linguistics;³⁰ it is therefore convenient to consider the last operation as a syntactical and no longer morphological factor.

From ornamentation, Gardin went on to the analysis of a more complex field: that of *iconography*. Here, as he notes, 'as compared to an axe or a pitcher, a picture cannot generally be adequately explained by the simple juxtaposition of all of the elements of its description'. The well-known problem is encountered anew of the distinction between subject and object: 'the man attacks the lion', or 'the lion attacks the man'. In order to express the different roles that a given element can play in a picture, Gardin adopted the method which, he states, is the most common in language: the *declension*. For instance, he distinguishes five 'cases': subject (coin bearing an engraving of a *tree*), object (a man cutting down a *tree*), qualificative (a man seated,

holding a *branch*), instrumental (a hero slaying a lion with a *branch*), locative (a woman seated on a *tree*).³¹

It is also necessary to express *actions*: in so far as the field of iconography is concerned, Gardin considers that merely two types of action need to be expressed, performed to the detriment of the object or not, negative or positive, or, as he states, 'black' or 'white'—good being simply defined as a neutral 'non-bad'. The exact sense of these terms is sufficiently defined by the context.

The *topographic connexions* form the object of a small number of diagrams: procession or simple juxtaposition of beings, with no bond between them (type . . . S . . . S . . . S . . . , S here being a 'subject'); the remainder in the form of a chain (type S → S → O, O being an 'object'), opposition of objects (O ← S → O), opposition of subjects (S ← O → S).

It remains to establish a *vocabulary* of representations. These are divided into a certain number of categories, ten in all, established in terms of the subjects actually found in the images studied (Greek coins, Oriental cylinders) and forming two large groups: *animated beings* (persons; animals; hybrids) and *inanimate objects* (garments; buildings; furniture; containers; instruments; weapons; emblems; nature—sky, earth, plants; ornaments and and 'indeterminates'). The *details* of human beings are specified (e.g., beard, hair), of animals (e.g., horns), of garments (style, cut, surface); the *attitudes* (posture, gestures) and the *viewing angle* of animated beings. The formal analysis of the subjects represented is not developed, which are considered as 'entities', each defined by a drawing and identified by a conventional symbol, and distributed among types at a level of generalization which permits generic research (e.g., huts, dwellings, temples, each type including some 'variants').³²

Gardin and his collaborators and associated researchers in their most recent research have extended the application of these methods of conceptual analysis to *texts*: Mesopotamian tablets, the Koran, mythical tales (and more especially the myths of the Pueblo Indians).

With regard to the *Mesopotamian tablets* the procedures used were somewhat similar to those outlined above concerning iconography. The 'categories' of the vocabulary are naturally adapted to the field under study. Here again we find humans, animals, and 'mixed' beings; the parts of the body, garments, instruments, constructions, to which are added art and transport. The 'nature' part is differently articulated; cosmography, fire, organic life, plant life, soils, minerals. There are categories of location and topography. A 'psychism' field appears, and social aspects are apparently being developed: status, institutions, kinship; military, legal, economic; religion, magic; sciences

(descriptive, metrological, astronomical, mathematical, divining, medical); culture.

'Verbs', on the other hand, are much more extensively differentiated; they include two additional indications: symbol ← for reflexive voice; symbol < for the past. They form the subject of a double-entry table which is reproduced below:

	<i>f</i>	<i>s</i>	<i>p</i>	<i>m</i>
		3 <i>Status</i>	4 <i>Property</i>	5 <i>Movement</i>
1. <i>Physical</i>	<i>fa</i> : manufacture, construct create; action giving birth to a being from the point of view of man	<i>sa</i> : institute, establish, giving birth to abstract being or function	<i>pa</i> : purchase sell <i>pam</i> : persons <i>paz</i> : property	<i>ma</i> : travel
	<i>fe</i> : maintain, treat, transform	<i>se</i> ¹ : organize, maintain, reform <i>se</i> ² : pardon, forgive, free a being from servitude	<i>pe</i> : give <i>pem</i> : persons <i>pez</i> : property	<i>me</i> : displace
	<i>fi</i> (opposite of <i>fe</i>): <i>fix</i> : natural disorder <i>fiz</i> : disorder, irregularity, ill treatment of things <i>fim</i> : same, of living beings	<i>si</i> ¹ : disorganize <i>si</i> ² : revolt <i>si</i> ³ : piracy <i>si</i> ⁴ : flight	<i>pim</i> : seize, hold, appropriate <i>piz</i> : property of others <i>pix</i> : goods of no ownership	<i>mi</i> : message
	<i>fo</i> : <i>fiz</i> : destroy property <i>fom</i> ¹ : kill persons <i>fom</i> ² : condemn to death	<i>so</i> : abolish, cancel		<i>mo</i> : transport of goods
2. <i>Professional</i>	<i>fu</i> ¹ : raw materials, production <i>fu</i> ² : secondary materials, transformation <i>fu</i> ³ : minor and temporary materials, <i>fu</i> ⁴ : liberal professions			

The *conceptual analysis of the Koran*³³ drew from the aggregate of *suras* less than 500 concepts, distributed among six large groups of categories:

Ontology

Chronology: beginning, past, present, end, beyond (there is no so-called future)

Natural medium: cosmography, reigns

Anthropology: biology, psychology, persons, society, politics, law, economics, culture, equipment

Philosophy-religion: notions, manifestations, rites, behaviour

Ethics: condition, conduct, universals, virtues, vices (deceit)

As in the case of the Mesopotamian tablets, these categories are not the product of an *a priori* classification, but the result of an empirical study of notions as they appear in the original texts and according to the bonds found among them, a result somewhat similar to those of the analyses performed by J. Trier or of G. Matoré,³⁴ on the one hand, and on the other to the 'classification systems by periods' indicated by Lund and Taub.³⁵

The last of the studies undertaken under the impulse of Gardin that we must consider here is that concerning *myths*.³⁶ The 'component units' of a mythical tale (phrases describing a certain manner of being of an Agent or of a Predicate) are: 'indexed and grouped by "categories" in a first attempt at conceptual systematization (e.g., family relationships, provisioning of food, etc.). No fixed rule should guide the selection of the categories other than a principle of simplicity: the "best" categories are those which constitute together the most economical system of reference in order to provide the greatest number of sequences, and ultimately a greater number of tales'.

The myth is then transcribed in the 'analytical tables', the horizontal lines of which represent the 'successive stages of the scenario' and the columns the above categories, arranged by 'homologous groups' (family relationships, 'space-time' etc.).

We shall not dwell on the following stages of the analysis, which aims at clarifying the structure of the myth, and replacing it within the anthropological framework of the latter, in order to permit a comparison among structural models of myths of a similar nature, or of a given cultural area, or a given period, to constitute finally a 'universal grammar' of myths: 'catalogue of fundamental representations' (lexicon of 'mythemes'), 'inventory of abstract methods of combination (morphology) or of transformation (syntax)'.³⁷

We have expatiated at rather great length on the work undertaken by Gardin, believing that it deserved attention well beyond the field of study (archaeology in a very broad sense) to which it refers. His method, indeed, seems to have very broad

implications for the entire field of codification for the retrieval of information—as soon as this begins to extend beyond the purely ‘bibliographical’ and superficial stage, towards the level of analysis in ‘depth’ of the conceptual content of documents.

Since 1960 Gardin has also been the organizer of the Groupe d’Études Documentaires pour les Sciences Humaines of the Association Marc Bloch, formed under the auspices of the École Pratique des Hautes Études and which has been entrusted under contract by Euratom with ‘establishing a general system of logical reduction’ of language, suitable for specialized literature in the humanities,³⁸ but with a more general purpose in mind: The humanities, or human sciences, were in fact chosen mainly because of the poorly formalized character of their current modes of expression, which is close to natural language; the problems of linguistic reduction present themselves here therefore in a somewhat exemplary way, and one may hope that the solutions proposed will have a general application for other projects of mechanized information retrieval in some particular field of science or technology.³⁹

From the point of view of our subject these new studies by Gardin appear to be extremely promising. They consist on the one hand of the establishment of a classification by conceptual categories of the ‘units of the dictionary’, according to a principle of ‘unique but multivalent’ opposition between two groups of ideas, ‘P’ and ‘Q’, contrasted in pairs—inorganic/organic, inert/living, animal/human, static/dynamic, physiological/psychological, objective/subjective, fixed (stable)/temporary, innate/acquired, natural/cultural—which makes it possible to arrive at a continuous semantic series, from the entities with dominant ‘P’ characteristics (physical elements) to those in which the ‘Q’ characteristics (society) are the more marked.⁴⁰ On the other hand, and this is most important, Gardin has made here a study of the categories of relationships that aims at simplifying their analysis. He first distinguishes three categories—function, process, way of reasoning⁴¹—and then retains only two of these: ‘Mode’, i.e., various, ways of relating observed facts one to another,⁴² and ‘function’. The indications of ‘mode’ are valid in principle for all terms arising from the analysis of a work; the indications of ‘function’, on the other hand, are ‘applicable in principle to pairs of terms whose reciprocal role they define’.⁴³

For the ‘functions’ Gardin distinguishes two levels of analysis: at the first, the general or ‘thematic’ level, he observes the presence—in texts which he has studied—of a central subject or ‘main variable’, of independent variables (factors which may be in relation to the central subject), of ‘parameters’, i.e., beings, space, time.⁴⁴

At the second level, which he calls 'specific', there are 'syntagmatic units' (or 'couplings'), in which the function of a term is defined in relation to a single particular term of the context.⁴⁵ At this level Gardin estimates that four relationships are enough "to record the chief "situations" observed, for every term or class of the dictionary in its relationship to any other term or class'.⁴⁶ These relationships are as follows:

Predicative, attaching to a term indicating an element or entity practically autonomous as an object of study, an essentially dependent property which describes its state, quality or function, i.e. predicate.

Consecutive, or, of causality, finality, etc., joining two elements of which the presence or action of one affects the presence, state or status of the other. This relationship can assume various meanings, according to the nature of the two elements concerned: the opposition of active and passive, a genetic relation, causal relation, conditional relation, functional or factorial relation, bond of finality, mediatory relation.

Associative, defined, in opposition to the consecutive relationship, as joining two elements of which the mention of one implies the simultaneous mention of the other. This relationship also assumes different meanings according to the context: relation of the part to the whole, of place, of hierarchy, of appurtenance, of specification, semiological relation. It is parallel to the predicative relationship; the difference being that in this an element is considered as dependent ('predicate'), whereas the associative relationship joins two elements both considered—in the dictionary—as being independent.

Comparative, indicating an extrinsic bond between two elements (independent, as for the associative relationship) of the dictionary: the author weighing up any two characteristics, in order to differentiate them as to their nature or their function in the same context.

These four relationships form a system, which Gardin tabulates as follows:⁴⁷

Intrinsic relations

static

with 1 or 2 predicates: predicative

between 2 elements: associative

dynamic

consecutive

Extrinsic relations:

comparative

Gardin's system of relationships is striking by the great simplicity which its author, very consciously, has given it, and which is in contrast to the comparatively much greater complexity of other systems (for example those of Pagès in France, of Ruvin-

schii at Euratom, of Perry-Kent or of Newman in the USA); the system is closer to the earlier attempts of Farradane (see p. 93) or of Kervégant (see p. 20, reference 8), but constitutes, however, an original construction, whose advantage over its closest two predecessors is that it is based on the analysis of a very great number of concrete cases without any previous prejudice.⁴⁸

Another peculiarity of Gardin's system of relationships is their 'mode of application'—if we may be allowed this expression. It proposes in fact not to apply the relationships at the level of the concrete, specific terms of the dictionary, but to make them operate between the more general, more abstract categories. Thus, in the example 'comparative statistics of schizophrenia in rural and town populations in Protestant countries', the consecutive relationship would not be applied between 'schizophrenia' on the one side and 'town' or 'country' on the other, but between generic terms such as 'residence' and 'illness'.⁴⁹ The objective aimed at here is a simplification, and therefore an economy, both in memorizing the data and in the retrospective research itself.⁵⁰

219 *The Isambert classification system for the sociology of religions*

As a last example of contemporary French research, we will mention very briefly the 'classifying categories' elaborated by François Isambert (Centre d'Études Sociologiques du CNRS) 'for a systematic bibliography on the sociology of religions'.⁵¹

Isambert recognizes (p. 147) 'the impossibility of a single classifying principle' and distinguishes three 'perspectives': reflexive standpoint, global presentations, analyses of structures and relationships, these latter being distributed according to analyses of the internal religious medium, then of relationships of the religious body with the external medium. He discusses (p. 148) the distinction between morphology, physiology, and dynamics, and the divisions of these three categories. His observations appear to be particularly useful, for they are those of a specialist seeking to find, as he expresses it, by 'groping', by a 'manipulation of stacks of cards', the 'various joints' of the 'bibliographic material'. We have strayed far from the application *a priori* of 'general categories', such as those of Ranganathan, but certain general notions would nevertheless be found (such, for example, as those of evolution, under Isambert's heading 'Vitality: regression and expansion'.)

22 Netherlands

Holland is probably one of the countries where, during the last thirty years, the most extensive research has been carried out in

the field of classification; this effort, however, has been practically entirely directed toward the development of the UDC. Consequently, we shall have only a few studies to record here.

The first to be noted is that made by C. Groeneveld (Royal Dutch, Amsterdam). In his report to the sixteenth Conference of the FID in Paris in 1946,⁵² he brought out clearly what he termed the 'first principle' of classification: coherence (or logic) ('consistency')—according to which a certain genus can be divided into species only on the basis of one single well-defined characteristic (standpoint) at a time. This, it is true, makes 'good sense', but it is rarely put into practice, even by authors of the best 'faceted' classification systems. Groeneveld, a chemist, followed closely the example set by his discipline, and demonstrated in several tables or schedules various types of classification of organic compounds according to various standpoints clearly defined and separated: composition, structure, properties, applications.

Thereafter, it would seem appropriate to recall the experiments conducted by the Dutch Patent Office (Octrooiraad) during the last few years under the supervision and authority of Dr. C. J. de Haan, President of the Patent Office—the great importance of which is recognized in Europe. As early as 1930, Octrooiraad was using in certain fields 'combined lists' bringing out, by columns, the presence or absence of various characteristics of the object of a patent bearing a given number—identical in their principle to various 'uniterm' systems.⁵³ However, it was only during the last two years that research has been undertaken with a view to mechanizing the documentary operations required for the preliminary examination of patents; the field of experimentation selected was that of carburettors.⁵⁴ A coding system in four sections was established, using the 'facet' method or categories, among which the basic elements of the description of this type of apparatus were distributed:

- A Principal aspect
- B Details of devices
- C Components of flux
- D Relationships (relative positions of the devices or the passages of flux; connexions between devices; junctions of passages of flux . . .)
 - Control system
 - Auxiliary functions of the devices

The most interesting aspect is the existence of a 'relationships' category, recognized as necessary for the description of carburettors. In order to permit the assignment of relationships in the coding system, as well as systems of control and the auxiliary func-

tions of a specified device attributed to the latter more or less equivocally, the Dutch Patent Office makes use of an interrelation procedure by 'interfixes', such as that used by the US Patent Office.⁵⁵

23 Germany

The International Automatic Documentation in Action (ADIA) Conference organized under the auspices of the Gmelin Institut and of the Deutsche Gesellschaft für Dokumentation at Frankfurt, from 9 to 12 June 1959, has given a rather complete outline of the research conducted in (Federal) Germany on modern methods of information retrieval; fairly detailed 'Vorberichte' (reports) have been published concerning the latter, which make it possible for us to dispense with long accounts in this chapter.

To date, the most advanced studies with a view to mechanizing documentation have been made at the Gmelin Institut under the impulse of Professor Erich Pietsch, who has also been presiding since 1951 over the Commission for the mechanization of documentation of the Deutsche Gesellschaft für Dokumentation,⁵⁶ as well as the FID/S (Selection) Committee of the Fédération Internationale de Documentation. Various methods have been tried out at the Institut since 1946.⁵⁷ A classification code was published in 1957 for organic chemistry,⁵⁸ and others for minerals, platinum compounds, the application of radiation to food industries.⁵⁹

The 'Systematik der Sachverhalte', (systematic presentation of contents) is a classification scheme, first, for 'matter' (chemical substances) based on the principal element (Leitelement) of each substance; this is later characterized by its 'state' (solid, etc., as well as notions such as adsorbed, activity, catalyser, etc.). Then follow the characteristics concerning the 'processes' or 'properties': physiological effects, uses, economic factors (production, consumption, etc.), extraction and treatment of minerals, analysis... preparation, treatment, defects... physical properties (general, then those of the core, of the atom, of the ion element, and of the molecule), crystallographic... electrochemical behaviour, chemical behaviour (and corrosion). For further details, alphabetical listing is sometimes used (e.g., under 30 formation, and 31 preparation). There are no special methods for expressing relationships; the subject is limited to inorganic chemistry.

The research conducted by Martin Scheele is of a more individual character, although from 1953 to 1956 he had directed the limnological documentation service created on his own initiative at Hannoversch-Münden, but later abolished upon the retirement of Professor Thienemann. In his book of 1954,⁶⁰ he

covered coding problems briefly, recommending, in general, a basic binary division into a 'system of objects' and a 'system of questions', and examined also the relationships between 'organ' and 'function' notions.

In 1955, he presented to the FID/S Committee, at the conference in Brussels, a report to which was attached an appendix defining the 'system of objects', divided into 'existential levels', bearing a striking resemblance to Glangeaud's scalariform classification system.⁶¹

In an article which also appeared in 1955, he developed further the 'system of questions'.⁶² Finally, he gave examples of a practical application of his method in tables (prepared with the help of IBM cards) of the *Archiv für Hydrobiologie*.⁶³

Also worthy of mention is the research conducted by the Arbeitsausschuss Medizin of the DGD, which intends, according to Dr. O. Nacke, 'creating a common "machine language" for medicine'.⁶⁴

24 Great Britain

As we have mentioned earlier, the principal centre of research on classification in England since 1952 has been the Classification Research Group, of which we have elsewhere indicated the origin, and upon which Ranganathan has exercised a profound influence. Before examining a few of the special classification systems elaborated by the members of the CRG, we must, however, recall briefly the previous research conducted by Holmstrom and Brisch.

241 J. E. Holmstrom

No doubt inspired by Kaiser (see p. 44) J. Edwin Holmstrom has advocated a classification system which he has named 'classification under ramified keywords',⁶⁵ and which was adapted by him or by other documentation specialists to various technical fields. It is, in effect, an alphabetical classification system by subject headings, but with this interesting peculiarity, that between the word chosen as the heading and the adjectives which explain the subject further, a letter (mnemonic in English) then proceeds to define the type of relationship between the main concept and the one or more secondary concepts. The types of relationship noted are the following:

- A action
- C corporate name
- E equipment

- G geographical name
- P personal name
- Q quality
- S substance
- T title of a publication on what is described

Holmstrom also mentions a method which he names 'grammatically inflected keywords' (p. 81-3) used by him in the 'coding' of the contents of military files during the last war; to an abbreviation of the verb indicating the action are added inflexions (from the English language), -ing for current action, -d for past action, -able for possible action, as well as the question mark, if the message implies a question, or the exclamation mark, if the sense of the message is an order; the addition of an N (or a stroke through the verb) denoting, on the other hand, a 'reverse sense'. In this method the verbs 'are selected from a standardized list, account being taken of the meaning, not the terms, of the "message" in question'.

242 *E. G. Brisch*

Brisch has spread in England, and later in the United States, a classification method of *objects* (materials, pieces, etc.) used in industry, in order to facilitate their standardization with a view to eliminating unnecessary duplications.⁶⁸ We believe it will serve a useful purpose to mention this method here, since its principle lies in a description of such objects according to their elementary characteristics, but selecting the objects according to the particular needs of the user, and removing for him any useless characteristics. It will be recognized that there is here something similar—in an entirely different field and for entirely different needs—to several of the methods which we have examined, and especially to Gardin's method.

The elementary characteristics are grouped by categories: form, nature (raw materials), function, dimensions, etc. Here again, we find a procedure already frequently encountered in the course of the present report.

243 *J. E. L. Farradane*

In three articles published in 1950, 1952 and 1955 Farradane tried to solve the problem of relationships in an original way. On the basis of the work of certain psychologists (Vinacke, Isaacs and Miller), he distinguishes nine relationships, obtained by correlating two series of characteristics: the first, concerned with the 'temporal' nature of the relationship, which can be (from this point of view) non-temporal ('non-time'), temporary, or fixed (permanent);

the second, concerned with the distinctness with which the relation is perceived. From this second point of view the relationship can be simply one of co-existence ('concurrent'), the two concepts being present together in consciousness without their mutual relation being otherwise defined than by this co-existence; it can be 'non-distinct', the two concepts being in this case directly tied one to the other, but without their method of attachment being known precisely; finally, the relationship may be 'distinct', the two concepts then being clear and explicit, but still definitely related.

It is possible to represent these nine relationships in the form of a table, with double entry, in which the squares contain the terms which Farradane himself gave to his scheme of relationships:

<i>Concurrent</i>	<i>Non-time</i> co-presence of 2 otherwise unrelated concepts	<i>Temporary</i> comparison, or relation agent/activity	<i>Fixed (permanent</i> association (for) also subjective properties
<i>Non-distinct</i>	equivalence (synonymity)	'dimensional' relation: properties derived from environment	belonging to (genitive)
<i>Distinct</i>	distinction from, or substitution for, imitation	action upon	causation or functional dependence

Farradane uses, to indicate each relationship, a symbol composed of two signs, which he calls the 'operator'; he also uses square brackets to enable multiple relationships of one concept to be written in linear form.⁶⁷

244/246 *The Classification Research Group*

Among the twenty or so special classifications established by various participants in the CRG, we will retain here those evolved by Foskett, Vickery and Miss Kyle. The names of D. W. Langridge⁶⁸ and E. J. Coates could also be mentioned; the latter established the classification system for the *British Catalogue of Music*.⁶⁹

244 *D. J. Foskett*

D. J. Foskett is the author of several special 'faceted' classification

systems of which three, at least, have been published. He was, with A. J. Wells, one of the first to introduce Ranganathan's ideas into England.⁷⁰

The Metal Box Company's classification system⁷¹ comprises six 'facets' (categories), of which four relate to the manufacture of boxes (products, parts, materials, operations) and two for packing and crating (packed and crated products—and material condition of the latter; processes). 'Various common subdivisions' are also added: research, development, instruments, control, special operations (welding, stamping, etc.).

The classification established for the food industries⁷² being an extension of the CC, utilizes the latter's categories, but refines their meaning: 'personality' becomes products; 'material' becomes parts, on the one hand, and materials, on the other; 'energy' becomes operations.⁷³

The most important of the classifications compiled by Foskett is the one on health and occupational safety, of which the schedules were first published as an appendix to the proceedings of the Dorking Conference,⁷⁴ then continued, modified and completed to serve as a classification for the International Information Centre for Occupational Safety and Health in Geneva.⁷⁵ The final schedule is as follows:

B	Physical agents and natural phenomena
C-G	Substances
H-J	Premises, equipment, processes and operations
K	Organization of labour and industrial structure
L	Fire and explosions
M-N	Pathology
P	Physiology and psychology
Q	Research techniques
R	Medical prevention and treatment
S	Techniques of safety and health
T	Equipment for individual protection
V	Organization of safety and health
W	Categories of persons
X	Industries
Y	Special aspects
Z	Generalia

There is also a certain number of common subdivisions, some general, others applicable specially to the field of medicine, and divisions for countries and international organizations. In short, divisions B to K are the divisions of a main category 'occupational risks', L and M-N form part of a category 'consequences of these risks', P-Q covers 'research techniques', R to V the 'protective and corrective measures', W to X the 'locations' of the phenomena;⁷⁶ but one sees that it is difficult to connect

these categories, which are specific to the field concerned, with general categories such as those of Ranganathan. The system is decidedly pragmatic.⁷⁷

It is not our intention to examine this classification in detail. Let us simply point out that complex subjects are not always formed by the combination of simple numbers. For instance there is no number for measurement in general, and one finds the various categories of measurements specified by direct divisions of the facet 'Q': the measurement of pressure at *Qac* (pressure, however, being *Bbg*), measurement of temperature at *Qaf* (temperature is *Bef*), the measurement of light at *Qcy* (lighting is *Bql* and *Smt*), etc. The measurement of gas in the blood is at *Qsg*, but there is no heading for blood in facet 'P' physiology; blood on the other hand, is present in pathology at *Mtz* diseases of the blood in general (one finds in physiology a heading *Pah*, humoral and cardio-circulatory modifications in the course of muscular work).

One of the curious characteristics—but not one of the more practical—of this classification is that the specific precedes the general, and that the long numbers are classed before the short numbers (such as *Gtz Skz* struggle against dust, before *Gtz* dust).

245 B. C. Vickery

In his book, *Classification and Indexing in Science* (p. 146-56), Vickery provides three diagrams of special classification schemes: two compiled by himself and the third in co-operation with J. Roland Smith.

The first—and the only one published in detail—relates to soil earth science; it consists of eight categories:

- Kinds of soil (subdivided according to various points of view)
- Structure
- Constituents (including organisms)
- Properties
- Natural processes
- Operations on soils (including the substances used as fertilizers)
- Laboratory techniques
- Generalities

The latter heading is of the greater interest from our point of view. It is divided as follows:

- Properties of qualities and of processes (intensity, stability, speed, variation, limits, mechanism, cause)
- Apparatus, equipment, instruments (curiously enough, under

this heading are found essentially actions: construction, tests, maintenance, norms)

Operations on the processes (starting, acceleration, lag or delay, prevention, control, automatic operation)

Logical operations (observation, experiment, definition, hypothesis, deduction, generalization, theory, law, computation and formulae, model, comparison, classification, nomenclature, criticism)

In contrast to Foskett, who does not use any relationship symbols (his symbols combining directly among themselves), Vickery uses here an oblique stroke / signifying 'relationship' and more particularly 'influence' or 'effect'.

It will be observed that Vickery, in this practical scheme, is unfaithful to his theory of categories, at least on one particular point: in the 'operations on soil' category he incorporates 'substances' (fertilizers and other soil improvements).

The two remaining diagrams published in *CIS* consist in a grouping of terms found in the DA to DF classes of the Bliss classification system for astronomy, and a classification for nuclear reactors, proposed as an extension of the UDC.

The classification for astronomy distinguishes ten categories: celestial bodies; their parts; their systems; their properties; the properties of their systems, movement of the bodies; relations and interactions between them, operations (human); instruments.

The classification for reactors is in nine categories: theory and experiments; control and conduct; components; fuel, uses (in general); piles according to their purpose; piles according to their cooling; piles according to their fuel; piles according to the energy of neutrons.

Vickery has also published recently a guide to the use of the 'facet' method in specialized classifications.⁷⁸ It contains a chapter on 'fundamental categories', and this is completed by the following chapter on the 'order of combination of facets',⁷⁹ and a very brief passage on 'relational terms'.⁸⁰

246 *Miss Kyle*

Since April 1955 Barbara Kyle has worked on a classification of social sciences, to be used in the various bibliographies of the Comité International pour la Documentation des Sciences Sociales de l'Unesco. Various duplicated work sheets and drafts have been published since March 1956, the most important—from our present standpoint—being document V (April 1957) which gave a provisional table of the classification. The latter had previously been put to a restricted practical test (document Va, September 1957, and alphabetical index, document Vb); a grant from the

National Science Foundation will shortly make it possible to test it more completely.⁸¹

In contrast to the Pagès code, which was intended for the analysis of concepts contained in documents, the Kyle classification is a bibliographical classification, and therefore at a more superficial 'level'.

In the last stage of the classification which has come to our attention, that of an edition dated March 1961, her general table is as follows:

- (0) Size, dimensions, proportions, level of development, shape, colour, speed, values
 - (1)-(2) Chronology, time, historical periods
 - (3) Form or type of writing, form of treatment, form of presentation, circumstances of production
 - (4) Special subdivisions (methods of study) for schedule A
 - (5)-(9) Special subdivisions for FT Law
 - 0 Non-human personalities and physical entities
 - 1-8 Geographical divisions
 - 9 Environment, locality, region, ecology
-
- People
 - b* By age
 - c* By sex and (family) relationships
 - d* By other biological factors
 - f* Intelligence and skill
 - g* Psychological types
 - h* Races
 - j* Nationalities
 - k-m* Languages
 - n* Religions
 - p* Professions
 - q* Obtrusive groups by own behaviour
 - r* Groups according to their size and associations
 - s* Classes
 - t* Structural groups
 - v* Firms, commercial undertakings
 - x* The whole population
 - y* The State
 - z* As agents
-
- A Disciplines, fields of study, canonical classes
 - B Physical and psychological activities of man and his surroundings
 - C Communication
 - D Arts
 - F Family, personal and social relations and society
 - G Education
 - H Breakdowns in society and their remedies
 - J Management and labour

K	Trade and distribution
L	Finance
M	Agriculture and economic services
N	Industry
P	Politics
Q	Parties: voting and elections
R	Local government
S	Central government
T	Central administration
V	International relations
W	International organizations
X	United Nations
Y	Wars
Z	History

One can observe in the detail of this classification an effort to systematize certain general categories, notably the (O) divisions. The divisions referring to people, from *b* to *z*, are in an order which starts with their physical and more stable characteristics, and passes to their characteristics as members of social groups (and for these passing from the less to the more structural).

In a note of June 1957,⁸² Miss Kyle discussed the relationship between the general order of her classification and Ranganathan's categories. She writes: 'The arrangement of the main classes must vary according to the user or the observer. At the present stage, we are attempting to create classification systems for special categories of users (in my case, researchers in the social sciences), and this must be defined before discussing the order. It follows, therefore, that a preferential arrangement for the main classes for a general classification cannot be established by an arbitrary decision as to what represents the smallest common denominator of users' interests. . . . The order of the main classes in my tables is that prescribed by the interests of users.'

Fundamentally, this expresses rather well the 'philosophy' which forms, expressed or not, the basis of all the CRG classification schemes examined by us. These are truly specialized, very 'pragmatic' classification systems, from which it would appear to be very difficult to derive elements for the creation of a valid general classification: the result would be a mere juxtaposition of special points of view.

Furthermore, and following the Ranganathanian tradition, these constitute *bibliographical* classifications at a fairly superficial level of analysis of documents, for which, in actual fact, the most interesting questions (elementary concepts, relationships, 'syntax', etc.) do not arise, or do not arise frequently.

Finally, and here also Ranganathan's influence is discernible, not one of the schemes of the CRG is conceived for application

in conjunction with automatic selection: they are classification schemes intended for traditional catalogues (or, in the case of Miss Kyle's, for bibliographies printed in book form), containing alphabetical indexes of the 'chain-index' type.

The observations which precede are not intended, of course, to belittle the value of the CRG's specialized classification systems, in so far as their use is concerned in fields and with methods for which they were conceived. On the contrary, in this respect, these systems appear to be perfectly valid (and certainly superior to the traditional, purely 'linear', enumerative, specialized classification systems).

247 *J. Seymour*

At the Royal Aircraft Establishment at Farnborough, R. C. Wright and C. W. J. Wilson adapted a classification, evolved by J. Seymour for aerodynamics, for an experiment in the retrieval of documents by visual superimposition of cards; it will be found in their report to the International Conference on Scientific Information.⁸³ It bears some resemblance to those of the CRG; its 315 elementary headings are grouped in categories:

- Bibliographical characteristics
- Velocity
- Fluids studied
- 'Types' of aeroplanes
- Components: wing . . . control apparatus
- Geometry
- Aerodynamical point of view
- Operation
- ...
- Technique and test equipment

248 *P. R. P. Claridge*

At the DSIR Research Station on Low Temperatures in Biochemistry and Biophysics, at the University of Cambridge, Claridge has undertaken an experiment in automatic retrieval of information (no longer merely of documents) on the chemical components found in edible plants, using the Filmorex (Samain) selection equipment. The code used is divided into six categories:⁸⁴

- A Bibliographical data
- B Botany
- G Growth
- H Harvest
- P Properties
- C Chemical components

to which are added a few additional indications (type of study, etc.). The coding of the chemical compounds was inspired by that of Samain at the CNRS; a classification of plants, based on that of Willis for flowering plants, has been conceived especially (p. 387-91) for the purpose.

249 *W. S. Batten and G. M. Dyson*

We shall merely mention here the names of these two British chemists. The first, while engaged in patent research at Imperial Chemical Industries (around 1939), re-invented the visual card superimposition method, which originated in the patents of the American Taylor (1915) and the Englishman Soper (1920) (see p. 58 and note 42 of Chapter 1), and later rediscovered in France by Liber (1923). Batten described the method in a report to Aslib in 1947;⁸⁵ quite naturally he applied the procedure to a coding system by 'points of view' (e.g., for plastics: chemical nature of the material, manufacturing processes, fields of utilization, details of the invention).

Dyson elaborated one of the most well-known and used coding systems for the description of organic compounds; in various subsequent articles, he examined in a very interesting manner the general problem of coding in chemistry.⁸⁶

25 Canada

For some twenty-five years, Hans Selye was engaged in the elaboration of a 'system of symbolic stenography (SSS) for physiology and medicine', which was published in 1956; a second edition appeared in 1958.⁸⁷

Selye criticized the existing classification systems in the above field—especially that of the UDC—because of their 'extreme complexity of symbols for relatively simple subjects' (p. 9 of the 'Procedure Manual'); he advocated the use of a system of short mnemonic symbols, presenting a great deal of analogy to that of Tchakhotine (see p. 62).

Selye's coding system is based upon a distinction between 'targets' and 'agents', i.e., a principle to the effect that, for classification requirements, the 'target' is always considered first, the 'agents' which influence the latter coming second, such agents perhaps also appearing as targets in other 'branches of the arborescent system': 'in biology, a target may frequently be converted into a negative or positive agent, by a reduction or an increase of its functional activity' ('Procedure Manual', p. 21-2). A basic rule of the system is that 'the effect of a negative agent precedes the effect of a corresponding positive agent' ('Procedure Manual', p. 23).

The coding operation is performed diagrammatically, using brackets, by the specialist himself, particularly in the numerous cases where several agents are tested on several targets, either individually or by a combination of the action of various agents (e.g. medicines forming the basis of individual experiments, and subsequently in combination, see diagrams p. 35-8 of 'Procedure Manual'). The parentheses are used to define the circumstances: chronic administration of a medicine, interruption of treatment, etc. ('Procedure Manual', p. 39) or again to identify a reaction on a 'relatively independent unit inside the organism'; e.g., treatment by the adrenocorticotrophic hormone, ACTH, of a local inflammation of a joint (*Art*) caused by formaline (*Fo*) will be given the following symbol: '(*Art*←*Fo*)' ← ACTH ('Procedure Manual', p. 40).

Most frequently, the SSS symbols are abbreviations of normal words of English technical medical language: *Ov* for ovary, *Nr* for nervous system, *Csf* for cerebrospinal fluid, *U* for uterus, etc. Many complex concepts are analysed in their elements: e.g., milk is symbolized *Ma-sen*, *Ma* representing the mammary gland and *-sen* (suffix) secretion (however, lactation is *Lac*; *Ma-sen* is used when mammary gland secretion itself is the target). Suffixes are used systematically, moreover, for notions such as enzymes *-ase*, contraction *-c*, internal surfaces *-endo*, function *-f*, inducement *-i*, inflammation *-itis*, sugar *-ose*, degenerative diseases *-osis*, syndrome *-S*, transplantation or transfusion *-t* (*-t* = transplantation as a spontaneous illness), extirpation *-X*; partial extirpation *-Xp*. The *m* = medulla is added directly, without hyphen (e.g., *Osm* = bone marrow, *Adrm* = adrenal medulla; likewise, *ac* = acid, which is used only as a suffix). But numerous normal 'morphemes' are suffixed: *Cav* (cavity or space), for example, will produce with *Cer* (cerebrum) *Cer-Cav* = cerebral cavities, with *CNS* (central nervous system) *CNS-Cav* = cavities of the central nervous system and with *Ly* (lymph) *Ly-Cav* = lymphatic cavities in general.

The symbols are grouped in tables which comprise twenty main classes ('Order of precedence tables') plus eight 'static categories':

- Nervous system
- Endocrine and sex glands
- Lymphatic and haemopoietic system
- Cardiovascular system
- Renal and urinary system
- Respiratory system
- Gastro-intestinal system, salivary glands, hepatic tissue, and biliary system
- Locomotor system, bone tissue, articulations and muscles

Cutaneous system, including its subsidiaries

Sense organs

Diffuse tissue systems, tissues in general

Cytology

Regeneration, healing of wounds and inflammation (in general)

Infection (in general), microbes, plants and invertebrates

Reactions to venomous animals, insects and parasites

Tumours (in general)

Metabolism (including diets and metabolism of pharmaceutical products)

Immunity

Conditions: internal medium (excepting metabolism, including general adaptation and genetic syndromes, general deformities); external medium

Paramedical subjects (in alphabetical order)

Generalities (current journals, history, others)

Description of interventions on the 'targets': techniques transforming the target into a negative agent (removal, etc.); techniques transforming the target into a positive agent (transplantation, implantation); processes *in vitro*

Normal morphology

Chemistry

General physiology

Pathological anatomy

Clinical aspects of illness or disease

A series of special symbols is used: the principal ones are as follows:

- ← Action (symbol inserted between the target and the agent to indicate damage, the production of a condition, a curative effect, etc.)
- To combine two symbols in order to form a third
- / Separation of coded elements in static codifications
- : Relationship between two targets, or between two agents
- ↓ Decrease, deficiency, or hypofunction
- ↑ Increase, excess or hyperfunction
- ⌊ Anti-compounds (e.g., \overline{Hn} = antihistamines)
- < In. For the chemical content, e.g., $A < Adr$ (= adrenalin in the suprarenals); used also to indicate the method of administration of medicine, e.g., $\overline{AB} > Csf$ (= antibiotics administered through the cerebrospinal fluid)
- > the same as <
- “ ” Pseudo
- (-) Deprivation, interruption of a treatment
- : - Without (absence of a characteristic trait)
- * Inserted following the symbol of an element, indicates its radioactivity.

The usual symbols ♀ and ♂ for female and male are used in case of need (e.g., ♀ -*Acc*, secondary female sexual organs).

The chemical compounds forming a part of a pharmacological group are surmounted by a horizontal line; spontaneous clinical illnesses (as opposed to experimental or induced illnesses) are underlined (illness symbols, in all classes, are similar to those for functions, organs, etc. in question: e.g., *Nr* represents nervous system, *Nr* represents diseases of the nervous system). We have already mentioned the use of parentheses; double parentheses (()) are used to designate processes *in vitro*. A complete list of the symbols will be found on p. 132-4 of the 'Code'; on p. 53 of the 'Procedure Manual', Selye suggests additional mnemonic symbols, e.g., \uparrow increase above the initial level, \downarrow decrease below the initial level, $\uparrow \downarrow$ increase followed by decrease, (+) mortality.

Despite a somewhat different symbolization, this coding system bears some resemblance in many respects to that of Pagès. Like the latter, and in contrast to the classification systems evolved by the English CRG, it is intended for the analysis of the *content* of documents: it has moreover been used and, writes Selye, has established itself as 'indispensable' (p. 55) for the compilation of reference materials and summary reports, the latter incorporating 'the salient facts selected from the 350,000 documents' of the institute's library.

It is a *specialized language*,⁸⁸ perhaps too much so, but many of the procedures involved therein are capable of generalization; it is, perhaps, an empirical method, but it is based on the experience of one of the most eminent researchers in the field of experimental medicine. For that reason, it seemed necessary to describe it here in some detail, and, we hope, without betraying its purpose.

26 United States of America

This part of our study will be both, and contradictorily, the easiest and the most difficult. The easiest by reason of the abundance of basic materials concerning research undertaken in the USA with regard to coding systems, and the availability of bibliographies, none of which seems individually exhaustive, but which supplement each other;⁸⁹ the most difficult by reason, on the one hand, of the tremendous dispersion of these materials, relating to a great number of studies and research projects, which are difficult to synthesize, and on the other hand, because of the unavailability (in Europe, or at least in France and Italy) of a rather large number of doubtless interesting documents, but which have appeared only in the form of 'technical reports' with a limited or restricted distribution.

We shall make an attempt here at regrouping the various systems of which we have some knowledge, in more or less

historical order, at least for the principal ones among them, those least readily individualized being examined later in somewhat 'residual' classes.

260 *Up to 1945-46*

As early as 1936⁹⁰ mention was made in bibliographies of a few isolated articles on the application of punched cards in various experiments on information or document retrieval, but it was not much prior to 1945-46 that the number of studies on the subject began to rise, primarily in the field of chemistry. G. J. Cox, C. F. Bailey and R. S. Casey, especially, published in 1945 a report on the application of punched cards to a bibliography on writing inks.⁹¹ The same year, D. E. H. Frear published the first article concerning a code devised by him in 1942 for the classification of several thousands of chemical compounds studied from the standpoint of their insecticide and fungicide action;⁹² this code served as a basis for the research undertaken by the Chemical-Biological Coordination Center established on 1 July 1946, under the auspices of the National Academy of Sciences and the National Research Council.⁹³ The CBCC constituted one of the most remarkable experiments in mechanized documentation on a large scale within the last fifteen years; for various reasons, however, it was abolished in 1957.

The CBCC's chemical code is of no direct interest here; on the other hand, from the standpoint of the present report, the biological code must be described in a little more detail. It consists of four main categories:

Descriptive and qualificative field: physical state, experimental state, administrative standpoint, direction of action, general action.

Field of numerical data: dosage of components, time, duration of action.

Classification fields (taxonomic fields): experimental substances, organisms, pathology, organs (primary and secondary), tissues, specific actions

Field of interpretations (degree of effectiveness, etc.)

Table 5 of Wood's article contains the code for the 'direction of the action':

- 1 Increases, stimulates, facilitates, etc.
- 2 Decreases, depresses, reduces, etc.: partial but not total stoppage
- 3 Stops, blocks, abolishes, etc.: complete stoppage
- 7 Produces, induces, etc. stimulates within the meaning of provoking an action which is not yet under way; 1 is used when an action which has already taken place is increased.

The purpose of these 'verbs' is to specify the 'specific actions' of the 'taxonomic fields' (classification fields), which include, for example, increase and differentiation, impairment (Wood's tables 13 and 14).

Taking into account the limitations imposed by the limited memory of the system used (IBM cards), certain factors—especially physical factors—had not been taken into consideration; similarly, the classification of human beings did not descend below the level of the species.

In an entirely different field, that of 'behavioural sciences', a coded analysis of ethnological data had been undertaken, as early as before 1939, by the Human Relations Area Files (HRAF) at New Haven, a research organization with a Yale University Institute status, 'incorporated' in 1949. The HRAF has coded the contents of a considerable number of documents of ethnographical interest, each page of a document selected for inclusion in the catalogues being annotated to indicate which categories of information can be found therein, among the 710 headings of a code named 'Outline of Cultural Materials' (OCM) embracing the entire field of human activity, and to which specific cultures they relate, the latter being inventoried by means of a second code, 'Outline of World Cultures' (OWC) enumerating some 2,000 distinct cultural zones and periods.

These two codes are of interest here, as examples of empirical codes, devised by specialists, and of which a study and, eventually, the critique or adaptation could prove profitable for a (future) general coding system—if this should ever see the light of day. It would be of interest to compare the OCM categories with those devised by Miss Kyle and by Pagès; a number of similarities would be noted, as well as rather substantial differences, which are probably due to the level of analysis for which each of these codes was conceived.

The HRAF now envisages the use of the Luhmann auto-analysis method.⁹⁴

261 *Calvin N. Mooers*

We have already encountered the name of C. N. Mooers at the beginning of the present report (Note 1). He must be considered as one of the most outstanding theoreticians on retrieval of information in the United States. He first became known in 1947 for 'Zatocoding', a superimposed method of coding by random numbers.⁹⁵ When the use of this method in various specialized fields, particularly chemistry,⁹⁶ was developed, Mooers was led to define a certain number of notions ('descriptor', 'interlocking sets', 'filtering principles') and to devise a mathematical theory of

information retrieval.⁹⁷ This last field, in which he achieved fame, goes beyond the limits of the present report, but it was impossible to omit mentioning here this great fund of research, the interest of which cannot be underestimated.

Mooers' 'descriptor' method consists in devising for each user a 'dictionary of notions' answering his particular needs. Each descriptor (if we may be permitted to use this neologism, which corresponds fairly closely to that coined by Mooers) represents an idea or a concept, generally fairly broad, and carefully defined by a 'scope note', or a note explaining its precise use, its particular significance in the specific individual system of which it is a part. The 'filtering technique' consists in an examination, for each document, of the complete list of descriptors incorporated in the system, with a view to retaining those which may characterize the document; this technique is simplified by the use of a list which regroups the descriptors under a series of 'chapter headings' presented in the form of questions.⁹⁸

262 *The American Society for Metals and the work of J. W. Perry*

Although the research undertaken by Perry and his team from Western Reserve University exceeds the scope of the American Society for Metals (ASM), we shall cover them in their general aspects in the present paragraph, since their most important practical application, as far as methods of coding are concerned, was made within the framework of an American Society for Metals contract.

Following an article by A. G. Guy and A. H. Geisler, of the General Electric Company,⁹⁹ a joint committee of the ASM and the Special Libraries Association (SLA) was formed in the autumn of 1948 for the study of problems relating to the classification of metallurgical literature. This committee developed a classification system known as the ASM-SLA system, published in 1950, and which subsequently received international attention.¹⁰⁰ It was primarily conceived in terms of a method of selection by cards with pre-punched margins.¹⁰¹

From the standpoint of the present report, the ASM-SLA classification scheme does not offer any particularly outstanding features; it is a classification by specialists for specialists. It is presented in three parts: index of processes and properties; index of materials; and the common variable index. The first index is rather ill-named since, in addition to a series of classes for processes (B-L), structures (M-N), properties and tests (P-Q), it also contains some 'generalities' (e.g., A, hygiene and safety, research organizations, etc.), raw materials (B, with extraction and treatment of minerals), inspection and control (S), applications of

metals (T), a composite class for corrosion (R) which includes both the mechanism of corrosion, corrosive agents, preventive measures and tests, and a 'related fields' class (U) for physics, chemistry, mathematics, etc. In the Italian edition, the latter part is entitled 'Methods', but this heading does not seem any more adequate for its content.

The index of materials is divided into four groups: Main chemical elements (and their alloys); Other elements (and their alloys); Materials classified according to their properties and applications; Metallurgical products.

The common variable index relates to equipment and processes; the factors which influence processes (high or low temperatures, pressure, etc.), types of products, the foundry, defects, etc., plus divisions relating to the form of documents, languages, and places.

Furthermore, the system of cards with pre-punched margins, to the requirements of which the code was adapted, operates satisfactorily only for quantities of less than 10,000 documents,¹⁰² and the ASM was very soon led to seek more adequate methods. With this end in view, it signed in 1955 a research contract for 75,000 dollars with the Center for Documentation and Communication Research of Western Reserve University, for 'a study to extend over a period of five years to determine the facts'.

The Center for Documentation and Communication Research of Western Reserve University, established on the initiative of Jesse H. Shera, Director of the Library School of the university, was from the very outset directed by James W. Perry, who, as early as 1947, had shown himself to be one of the most earnest supporters of the use of methods of mechanical retrieval of documents, and had presided over the Punched Card Committee of the American Chemical Society.¹⁰³

Beginning in 1954, a series of articles by Perry and his collaborators was published in *American documentation*: these articles were later supplemented and brought together in book form in 1956.¹⁰⁴

The most interesting innovation resulting from the above study concerned a coding method based on a 'semantic analysis' of complex terms into 'individual terms', a method already outlined by Kent and Berry in June 1952 at the Symposium on Machine Techniques for Information Selection, Massachusetts Institute of Technology.¹⁰⁵ For that purpose, an extensive collection of terms—approximately 30,000—had been assembled, derived from many different sources¹⁰⁶ and 7,000 of these were analysed on marginally pre-punched cards, into five 'general classes' according to the type of entity or of concept to which they referred (processes; machines, apparatus, devices; materials,

substances; common variables, characteristics; abstract concepts), and, on the other hand into five categories according to the 'general field' to which each term was related: chemistry, physics, mechanics, biology, science and technology in general.¹⁰⁷

The 'key' chapter of *Machine literature searching* was Chapter XI; 'Construction of machine language', which appeared for the first time in the book (p. 72-90), and outlined the breakdown into 'semantic factors'.¹⁰⁸ Strictly speaking this method cannot be considered as an innovation. It refers back at least to Leibniz, who had spent a lifetime working on its application, beginning in 1666 (*De arte combinatoria*) or even perhaps earlier.¹⁰⁹ On the other hand, among the documentalists, Cordonnier had rediscovered it in 1943, and had outlined it rather clearly;¹¹⁰ Pagès, on his side, had examined at length the same problems in 1948, in a more specific manner;¹¹¹ in the USA itself, the 'interlocking sets of descriptors delineating the idea structure' technique outlined by Mooers in 1951, and subsequently in 1955,¹¹²—and already but into practice by him as early as 1948—and the 'new method' presented by Luhn in 1951 'of characterizing a topic by a set of identifying elements or criteria'¹¹³ were identical. We draw attention to these 'anteriorities', not in order to underrate the work performed by the various researchers or teams of researchers—who, in most cases, truly believed that they had discovered a 'new method'—but to persuade them, rather than to advocate unilaterally any one 'exclusive' process, to agree that they are all engaged in work on common basic principles, whatever may be the differences (at times very minor) in the coding method or the particular type of machine adopted. . . .¹¹⁴

The basic symbolization procedure adopted was not an original one either: it used the 'self-demarkating code words' of three or four letters suggested by Luhn in 1953, with a few adaptations.¹¹⁵

There were, on the other hand, a certain number of features peculiar to the method developed by Perry, Kent, and Berry, and these we shall now review:

1. The distinction between 'analytic relationships' and 'synthetic relationships', the first of these expressed by a variable letter (the second) within a code-word of four letters indicating a 'semantic factor', the second being assigned special code-words of three letters.

It is somewhat difficult to understand clearly the difference between these two types of relationships.¹¹⁶ The authors explain that they indicate as analytical the relationships between concepts whose aggregate constitutes the machine language. In theory, their number could be extremely high (equal to the number of

pairs of concepts which could be constituted by taking the concepts of the machine language adopted two by two), but a simplified set only will be used, the definition and the selection of 'analytic' relationships retained being governed by practical considerations, 'the most important [of which] is their effectiveness as a means of defining and conducting selection operations, in order to separate adequately the documents of interest in connexion with a specific problem or to a request for specific information, and those of no interest in this connexion' (p. 78, 2nd column).

As to 'synthetic' relationships, Perry, Kent and Berry apply this term to 'relationships observed, presented as empirical facts or deduced theoretically from the documents' (p. 79, 1st column). They add that 'a given relationship, when observed and registered for the first time, must be considered as empirical in nature, and therefore of a synthetic character; subsequently, however, the same relationship can become the basis upon which a new concept is defined, and the relationship in question can then be recognized as being analytical in character'.¹¹⁷

'The set of symbols used for analytic relationships is entirely different and distinct from that used for synthetic relationships. When considering the definition of synthetic relationships, it must be noted that their elaboration is primarily governed by considerations of effectiveness in terms of the objective to be reached, i.e., they must present the maximum number of advantages when automatic machines are used in retrieval and selection operations.'

Taking into account this last phrase, one observes within this categorization of relationships into 'analytics', on the one hand, and 'synthetics', on the other, less a reflection of such a distinction in the facts of scientific experimentation, than the specific requirements of a given method of symbolization for a certain kind of selection equipment. We do not believe that we should consider this separation as a valid one in general for a 'common language' of codes for the retrieval of information; in fact, it does not seem to have been adopted by other authors.¹¹⁸

2. The code-words are used to transform 'telegraphic style analyses' of documents into 'coded analyses' (encoded abstracts) containing all the essential data in the document, including numerical data, in relation to the necessary processes for retrieving the information in it. In this connexion the analysis must be divided into distinct statements, then into phrases, in order to separate certain parts of the final encoded analysis, with a view to avoiding any confusion which the machine might otherwise make between concepts mistakenly linked to each other ('cross talk'; see p. 105 of the book). This is accomplished by special punctuation symbols (|| and ꞑ). Other available symbols are, the & indicating the

beginning and the end of a 'specific symbolization', that is, a concept (geographic, chemical element) which is not coded according to the code's normal language (its role is somewhat similar to that of the 'analytical prefix' of Pagès (see above); and the / which indicates a quantitative specification, or the name of an author, or a document reference.¹¹⁹

3. A limited number of hierarchical classifications are used: on the one hand for geographical locations, with a three- or four-level hierarchical code (continent or part of a continent, large region, nation, city), and on the other hand for classes of objects or of concepts indicated by one or more code words, but subsequently subdivided by means of enumerative divisions into a decimal notation.¹²⁰

In 1958, Perry and Kent published the 'Semantic code dictionary' which the readers of *Machine Literature Searching* had been awaiting expectantly.¹²¹

The 'encoded abstracts' method was presented in this work in much greater detail; especially in chapter 5, by Jessica Melton, entitled 'Procedures for preparation of abstracts for encoding' (but the current title 'Metallurgical abstracts for encoding' specifies that these are abstracts undertaken for the ASM) provides a complete outline of their technique which, in the terms of *Machine Literature Searching*, has been somewhat modified. The distinction between 'synthetic' and 'analytic' relationships has disappeared in this chapter, at least in the form in which it was presented in 1956, with its 'theoretical' justification. As far as synthetic relationships are concerned, mention is simply made that: 'the effective retrieval [of information] with the help of any low-cost electronic equipment which can be conceived in the present state of technological development, requires that the important aspects of the contents of documents be explicitly symbolized in a consistent manner. In particular, relationships must be indicated in a uniform manner, regardless of the various form of phrases used in current written language. With this objective in mind, a certain number of relationships have been defined more or less arbitrarily as being the most effective in recording the important aspects of the information contained in documents relating to metallurgy, for the identification of the latter. These relationships, of assistance in the retrieval of documents in this field, are indicated by a set of symbols called role indicators which operate as "syntactic signals" in automatic retrieval.' We have related this passage in its entirety (p. 76), inasmuch as it seems to indicate very clearly the empiricism of the Perry school. The selection of 'relationships' is based on

purely 'operational' considerations, in terms of a certain type of equipment at present available at low cost and the particular coding which seemed to be the most favourable for this type of equipment. The expression 'synthetic relationships' does not appear in this passage; one must return to pages 100-3 to find it; there again there is some insistence as to their arbitrary character. The role indicators no longer number 12 as indicated in the 1956 volume (it is true, as 'examples', p. 89), but now number 23, grouped into four large categories, materials, properties, processes, conditions, defined in pages 103-6, plus a 'miscellaneous' category. It is furthermore mentioned that: 'this does not represent an attempt at classification—except in the broadest sense of the word. The grouping of phenomena and of concepts in this manner is merely a handy rule-of-thumb method to prevent the creation of inconsistencies in the telegraphic analyses of a vast catalogue, as a result of the fortuitous manner in which phrases of an outline, or any phrase, are written, or the degree of detail of a specific analysis.'

Listed below are the 23 role indicators for metallurgy, in accordance with the 'glossary' (p. 128-46) ('prefixed for a term indicating . . .' is implied in each case):

1. Relationships for 'materials'*

- KOV 'material' for which the properties are indicated
 KEJ 'material' subjected to a process (this 'material' can be an object, e.g., a tanker subjected to corrosion, or a theoretical action)
 KUJ component: component of the above-mentioned 'material', or a subdivision to be isolated for the needs of research (e.g., the *plating* of zinc-plated copper piping)
 KQJ by means of (agent or instrument effecting a process, a testing technique or a function)
 KWJ product (end product of a process, 'material' detected or determined by chemical analysis, or intermediary product; in the latter case, KWJ is coupled with KEJ or KAJ)
 KAJ initial 'material' (subjected to a process or to manufacture, when the manufactured product is the factor of the greatest interest; little used in metallurgy, its utilization will probably be broadened in the field of organic syntheses)
 KAD machine or device of which the description and/or function form the main subject of an important part of an analysis
 KAG part of a machine or of a device

* See definition, p. 114.

2. *Relationships for properties*

- KWV indicated property of a 'material' and/or its components, not resulting from a determination included in the content of the document
- KUP specific property: property, measure, value, rate, etc., of which the determination or the process from which they result are presented in the document
- KAP property influenced by, depending from, varying with, or as a function of another factor indicated by KAL

3. *Relationships for the processes*

- KAM process, applied on a preceding 'material'; testing technique, observation or computation method, used to obtain observed or theoretical determinations as to the preceding 'material'; function of a machine or a device previously indicated; process which occurs in the 'material' under certain conditions
- KXM negative process: process or function which does not occur

4. *Relationships for conditions*

- KAH condition: circumstances or factors prevailing or accompanying the process, which may influence the latter or be necessary to its achievement

5. *Miscellaneous role indicators*

- KAL influenced by: material, process, etc., which affect a preceding property indicated by KAP
- KIS location: location where a process takes place, either inside a material or outside the latter.
- KWB direction from } location from which or toward
KWC direction toward } which a movement occurs
(a note indicates that these indicators are rarely used in practice, usually because other synthetic relationships, more commonly the object of document retrieval, can be employed, even when they contain a concept of direction as an element of the general idea)
- KAB scientific or technical field, or name of a theory
- KIT time: date or period
- KIG organism concerned
- KEP person concerned, or group of persons such as workers, syndicates, management, if they constitute an important aspect of the analysed data.

To the above 23 indicators, recognized as useful in the field of metallurgy, have been added 9 others in the course of an abstract encoding experiment conducted by Thomas H. Rees, Jr., in connexion with *New York Times* articles (p. 176-88):

- KXJ acting person or organism (agent of a process indicated in a subsequent phrase)

- KEB category of organism (type of organism: table II, p. 179, gives a list of 24 classes of organisms—aviation, automobile, . . . distribution . . . governmental organism, syndicate . . . research, naval construction, metallurgy, transportation)
- KIP position (within the organism), rank or title of the person considered
- KES subordinate organism (branch, agency department) when the name of the higher level organism is also mentioned
- KIN client or user: persons, organisms or classes of organisms for which another person is acting, another organism or another class of organisms
- KOJ source of the information: person, organism, etc. from which emanates the major part of the information contained in the article (may be similar to KXJ)
- KUG location indicated for the society or any other organism or group of individuals, when the following word is prefixed by KIG
- KAV specified quantity: price, value, productive capacity, salaries, production, consumption, and analogues, accompanied by numerical values
- KAC type of article (followed by another letter, to specify whether the article in question is a news item, a communiqué, a speech, a prediction, an opinion, an account, or a general study)

Furthermore, it became necessary to give slightly different meanings to seven of the indicators listed above: KIB has become organization concerned (as distinguished from KXJ); KEP indicates an individual person or a class of persons, and no longer a 'group of persons', which then becomes KIB, and it is the person concerned, KXJ being the person acting; KIG, place, is supplemented by KUG; KWC and KWB are prefixed to names of organizations or to countries, to indicate the direction of such activities as purchase, sale, import, export, loan, education, etc.; KAB, the field to which the article belongs, is used with an additional letter to otherwise specify the field in question (new applications, changes in prices or economic conditions, questions relating to labour, new installations, production, patents and inventions, research, or theory). KIT, date, is also re-defined as also covering the period for which a prediction is formulated.

The particular meaning attributed by the authors to different 'categories'—materials, properties, processes, conditions—should perhaps also be clarified here (pages 104-16):

materials: (1) entity or material means; (2) parts of a material entity for which a chemical composition could be indicated; (3) elemental particles of matter (in a special sense, since it also covers atoms and molecules); (4) forms and various types of

energy, such as electrical energy, ultrasonics, solar energy
properties: (1) measurable physical, chemical, or mechanical property of a 'material'; (2) values, rates, measures or performance (capacities) of a 'material'; (3) division of a 'material' for which the chemical composition could not be furnished (see above, under materials (2))

processes: (1) change or action which a material undergoes or that it should produce; (2) operation or theory (technique or method) used in treating a 'material'; (3) function of a machine or device

condition see the definition given under KAH (above); on page 106 of 'Tools . . .' it is indicated that KAL may be used with a term relating to one condition, although KAL, in the role indicators glossary, is classed in group 5 (miscellaneous)

We should mention, furthermore, that to these four categories of role indicators, Rees had added a fifth category: *relationships for organizations and persons*, where he had regrouped KXJ, KIB, KEP, KEB, KIP, KES, KIN, KOJ, KWC, and KWB.

The 'analytical relationships' in the 1958 volume, properly speaking, no longer seem to be specific, except by analogy with the 'infixes' of Semitic languages indicating grammatical relationships (p. 233), and by the examples which follow through page 235. The list of relationships which is given (p. 278-9) conforms to that already published in 1956, with the addition of a Z relationship, simulation. We now have, therefore, a set of ten infixes which, it should be recalled, are indicated by a 'variable' letter (the second within the words of four letters indicating 'semantic factors', the other three letters performing the role of 'roots': it is necessary throughout the following description of the use of the infix, to add at the beginning 'the coded word represents . . .' and at the end '. . . represented by the semantic factor'.

<i>Letter</i>	<i>Relationship</i>	<i>Use</i>
A	categoric	a member of the class . . .
E	intrinsic	that which is composed of that which exists . . .
I	inclusive	that which is a component of that which exists . . .
O	aggregate	something grouping (made up of) several members of a class . . .
U	productive	that which produces the object or is used for the action . . .
Q	affected	that which makes a use of, is determined or influenced by that which exists
W	instrumental	that which is produced by, acts upon, or upon which acts that which exists . . .
X	negative	that which is characterized in an important manner by the absence of that which is . . .

Y	attributive	that which possesses as one of its most important characteristics that which is . . .
Z	simulative	that which has certain properties of, but is not that which exists . . .

We shall not undertake here a detailed discussion of the above table of relationships, both 'synthetic' and 'analytic' in the Perry-Kent coding system; it would considerably exceed the bounds of the present report. We shall limit ourselves to a few observations:

1. The above is an 'overspecified' language, which the authors seem to have wanted to turn into an over-explicit instrument, where relationships are defined twice rather than once; this results in a weightier language than ordinary language, frequently leaving many implied relationships. This feature, however, may have been intended, in order to avoid ambiguities which the retrieval equipment may not have 'understood'. However, the system seems to overload the code and its symbolism excessively.

Let us take the following example (p. 129): the corrosion of a tanker. The PK code (abbreviation for Perry-Kent) expresses this notion as

,-KEJ.SAHP.003,-KAM.DATR.001

which means '*subjected to the process which will be hereinafter indicated: third member of the class of ships (S-HP) - process acting upon the object previously indicated: first member of the deterioration class (D-TR)*'.

The fact that the deterioration applies to the tanker, expressed in French by the small all-purpose particle *d'* (and in English simply by the apposition: tanker corrosion), is specified twice in PK, first by KEJ and then by KAM.

Another example: ironworks producing automobile and aeroplane parts (p. 138):

,-KAD.CUNS.30X.MAC.003,-KAM.CUNS.030,KAM.CUNS.
0.14,-KWJ.LAMN.028*HACL.002*SARR.001

which indicates: '*machine described: producing 30th operation (forging) of manufacture class and 3rd member of the class of machines so utilized, —function of the machine previously indicated: producing 30th operation (forging) of the manufacturing action, —function of the machine previously indicated: producing 14th operation (production) of the manufacturing action, —product: 28th member (part, component) of the "elements" (unit) class*

designated as follows: 2nd member of the vehicular class designated as follows: 1st member of the aeroplane class.'

[In the two preceding examples, we have replaced the ordinary words designating the semantic factors which we found on p. 129 and p. 138 (where the analyses are still 'telegraphic' and not yet 'coded') by the code-words indicated in the semantic dictionary; we hope we have not made any errors in the use of this code, which is rather complex for a non-initiate; the role indicators are underlined.]

It is probably unnecessary to calculate the 'coefficient of redundancy' of the PK code in the above example: it is, without doubt, far greater than that of any natural language.

2. The 'analytic relationships': corresponding in part to the distinction, in normal language, of 'parts of speech' and of time. Consequently, to insulate is SAPT.003, insulated SWPT.003, insulating (present participle) and insulation SUPT.003. But insulation (substantive) is much more complicated: SQPT.LWCT. PUTT.001—which means making use (Q) of the separation (S-PT), acting upon (W) electricity (L-CT), producing (U) a protection (P-TT)—the whole being the first division (001) of the complex class so formed.

The PK 'language' is, in this sense, a 'symbolization' language of roots—with vowel gradation—by reason of its 'analytic' relationships expressed as in ancient Indo-European or Semitic languages; it is 'inflexional' or 'coalescing' by reason of the prefixing of its 'synthetic' relationships, each of which strictly adheres to the word which it precedes and possesses no autonomy: from these two standpoints, it seems far removed from a language having pure syntactic relations such as the Chinese language, and it is somewhat astonishing to find John L. Melton comparing it to the PK code (p. 241).

However, in certain cases, it is not the 'analytic relationships' which are used for distinguishing the 'parts of speech', but a change in the other type of 'inflexion' consisting of the three decimal figures attached to each term of the code. For instance, class is RANG.004 (where R-NG is the root 'arrangement, disposition', and A the category 'analytical relationship'), and to class RANG.005, but classification is RUNG.005 (U being the 'productive' relationship). At times even the verb and the corresponding substantive are distributed among distinct roots: for instance 'to dissociate', SWPR.009 (S-PR = purification), with dissociator SUPR.028, but dissociation is at DUSM.007 (D-SM = dismounting, dismantling, explained as 'physical separation and by the force of physical objects', physical and forcible taking apart of physical objects).

The *semantic code* itself should now be examined. In its present state, as writes J. L. Melton (p. 232), it comprises 214 basic semantic factors; however, a reckoning of those given in the list in table 1 (p. 249-50) indicates 213 only.

The code is established in alphabetical order of 'etymons' (roots) in this list as in the 'Code-English Dictionary' (p. 747-964) which provides combinations of etymons with their analytical relationships and their decimalized divisions. Table II (p. 259-78), however, provides a 'general classification of the generic concepts represented by semantic factors' where they are 'arranged under various headings to indicate their comprehensive relationships. No attempt has been made to perform a logical analysis, and the same factor may appear under various categories'. The classification is as follows:

General concepts

Ideas

Relationships

Properties

Fields of activity

Forces

Classifications (according to the composition, nature, utilization)

Relationships

Economic

Physical

Social (basic social concepts; concepts affecting society; processes affecting society; factors affecting society)

States

Psychological

Requiring, implying an action

Physical (sensitive, perceptive states; conditional states; properties implying states)

Processes

Physical

General

Materials (creating a change; constructive, destructive)

Industrial and occupational

Substances

In general (generalized substances)

Specific: organic, inorganic

Preparations

Objects

General (natural; products)

Special (natural; products)

As far as we are concerned, we have not found this systematization particularly enlightening; but we are perhaps rebellious against any such 'general categorization', which seems to us to relate rather to scholastics.

The importance of the different semantic factors varies considerably. A few have only a very restricted number of divisions and combinations; for instance C-RD, raw materials has 2 divisions and 1 combination; R-LG, religion, 1 division only and 4 combinations; on the other hand, M-CH, device, occupies pages 835-48, and ten pages have been set aside for P-PR, physical properties.

One aspect is particularly striking at the outset: the *very great number of 'inflexional' divisions by 3-figure decimal numbers*. There are four pages of these under R-PR, relative term ('concepts which have no meaning other than in relation with something else'), spreading from 001 to 382 (with some gaps), in the alphabetical order of English terms from abnormal to wild (208), and, subsequently, in no apparent order. Under the said R-PR, one can find practically anything: analogous, sombre, domestic, accurate, standard of living, pure and impure, synthetic, module, classical, luxury.

R-PR is doubtless an extreme case; however, one finds under P-PR, physical properties, a long alphabetical enumeration (still with 3-figure decimal notation) extending to 086; under C-PR, chemical properties, up to 053; under D-CM, document, up to 062; up to 043 in G-PR, general properties 'properties of all types characteristic of non-specified objects': it is here that we find essence 039, and nature 040, with allotropy 034, competence 036, equilibrium 038, speed 042 and rapidity 041. Under M-PR, material properties ('minerals in general—generalized minerals and their properties') again we find three pages of decimal divisions, extending to 192, and which, furthermore, relate to many other things than minerals: we find, for instance, a large number of geometrical forms and space itself, in 091, with form in general 088, plus smell 008, and taste 086. Under R-CT, reaction (chemical), there are 159 enumerated reactions, practically without gaps.¹²²

Science, S-CN, has 52 direct decimal divisions for various fields of science, from alchemy to trigonometry; a very large number of other sciences, besides, are indicated by combinations.

One other general observation: the treatment of *oppositions and negations* is not uniform. In the case of equilibrium, GAPR. 038, a 'non-equilibrium' has been created by the use of the analytic relationship: GXPR.038; this method has been followed repeatedly in C-NG (change and reversal) perhaps because the title of the semantic factor itself evoked the idea. But isotropy

is PAPER.030 and anisotropy PAPER.004. Order is RANG.002, and disorder RAPER.051; true is RAPER.189 and false RAPER.070; good is RAPER.218 and bad RAPER.330. Generally speaking, opposites are often located in different classes. Cause is GANT.005 (G-NT = agent), and effect PARD.002 (P-RD = product); end is TARM.008 and origin BAGN.001 (start is at BAGN.006). Hot is a composite term: RAPER.RYHT.4X.001 (R-HT = heat, and RAHT.004 = temperature), but cold is a direct division of RAPER, RAPER.030.¹²³

It would be an arduous task to follow the *general notions* through the vast PK code. They are, indeed, quite scattered and it is rather difficult to reconstitute them in their entirety—perhaps even more so than in a UDC- or CC-type classification. Let us take a few samples only.

We have seen *space* and many notions in this field under M-PR, material properties. However, notions of form, of structure, etc., are found under several other semantic factors: first, under S-LT 'framework' (framework, skeleton, etc., the title is explained as 'structures around which things are arranged or constructed'), then in C-BL wire ('things in the shape of, or used as wires, cables, or filaments'), C-CP fill (explained as 'something inside something'), C-CR cover ('something placed on top of something') H-LL hole ('voids of all types or shapes'), L-CN location. Under the latter heading, explained as 'places and areas situated in space', are found direct divisions, such as 005 cemetery, 005 city, 022 hospital—followed by a whole series of composites, such as library LACN.DWCM.11X.001 (D-CM = document), magnetic field LACN.MWGN.001 (M-GN = magnetism), interior LACN.RYPR.006.

Time is under TMM.001, T-MM time being explained as 'process directly implying passage of time': two seasons only are found here, spring 021, and winter 022; day TIMM.010 and night TMM.005, lifetime TIMM.012; age is under TQMM.PASS.001 (P-SS = process), old in TYMM.RAPER.022 and youth in TYMM.RAPER.028 (between which are inserted obsolescence, permanency, previous, and sudden). Period is TIMM.011, periodical TWMM.RYGL.4X.001 (R-GL = control, RYGL.004 = regular), periodicity TWMM.PAPER.RYGL.4X.001 (P-PR = physical property).

Notions relating to *number*, to *quantity*, etc., are under C-LC computation, D-DT addition; number itself is LAMN.019, L-MN being 'unit: things used as reference marks to measure or indicate'; numerous is LYMN.RAPER.250X.006 (R-PR = relative term; RAPER.250 = multiple). *Measurement* is M-SR, 'a process implying a measure', there are here two and a half pages of composites. Mathematics, a direct division of science, is SACN.033,

algebra is SACN.006, statistics SACN.LQMN.001 (L-MN = unit).

We have examined this code in somewhat great detail, for it is up till now the most ambitious attempt of its kind to work out a method of coded analysis in a vast technical field. It is an undertaking of considerable breadth, which deserves respect. We do not have the means to judge the PK code according to criteria of effectiveness, which would be the only appropriate ones in view of its specific purpose, i.e. to permit the retrieval of information by means of relatively low cost selection equipment (although still much more costly than the visual superimposed cards used by Pagès and Gardin). It must be mentioned here that the Western Reserve University team of researchers has elaborated, in conjunction with this code, a 'strategy of research' which may be worthy of attention.¹²⁴

As regards its possibilities of adaptation outside its initial field and beyond its immediate objectives, the code's outlook seems, truthfully speaking, less certain; the analysis into the semantic factors forming the basis for its construction seem often to be of little utility elsewhere. Leroy and Braffort (Note CEA 278 already frequently cited, p. 4)—who were inspired by Jessica Melton's 'model', and notably by its diagrammatic presentations (p. 82-5 of 'Tools')—seem to believe that it is unreasonable to seek a 'universal language' and that it 'seems more rational to begin with a linguistic study of specific scientific fields, and to derive from these specific linguistic schemes, from which later a single scheme might be developed'. But is the code we have just studied intended to be a 'universal language'? This does not appear at all certain: as we have seen, it refers very frequently to the special requirements for the retrieval of documents in the special field of metallurgy. Perhaps one should seek elsewhere the reasons for the defects found in it when it is regarded from the 'common language' standpoint. We believe, in this connexion, that attention should be drawn to the characters of the symbolization adopted, the weightiness of which may be a contributing factor to many of the negative aspects of the results obtained and, in particular, also to the fact that fundamentally the semantic analysis of concepts is not extended very far. It will be noted that the PK code's 'composite words' never consist of more than four distinct semantic factors; this is probably insufficient, but in view of the uneconomical method of formation of elemental code-words¹²⁵ it was difficult to proceed farther.

On a more general plane, the code seems to have suffered somewhat from too empirical and too pragmatic a 'philosophy',

and also perhaps from a certain contempt for classification proper on the one hand, and from a certain failure to appreciate the results of linguistic research and modern philosophy, on the other.

As it stands, one can observe in it the evidence of a rather remarkable effort to leave the 'beaten path', and a detailed critical study of it will tempt the investigator to progress further.

263 *Research at the US Patent Office*

In 1900, the US Patent Office had available in its files approximately 1,200,000 documents for inspection by examiners in charge of the preliminary examination of applications for new patents; in 1955, that figure had reached approximately seven million. In order to facilitate the search for earlier patents in this important mass of documents, the Patent Office had little by little developed one of the most extensive classification systems in existence. It comprised, in 1948, 43,904 subclasses.¹²⁶ But this system, although added to as the need arose, no longer corresponds to present requirements.¹²⁷ The only possible solution to the Patent Office's problems of information retrieval seemed to lie in the application of automatic or mechanical methods; such was the main conclusion of the report of a special advisory committee, chaired by Vannevar Bush, in 1954.¹²⁸

A first experiment in that direction had been made as early as 1950¹²⁹ on class 260, carbon chemistry, and had been rather encouraging. Following the Bush report, the office of Research and Development undertook a series of basic studies on the conditions to be fulfilled in order to render the Patent Office's future automatic information retrieval system as effective as possible. To this date, 15 reports published in the series of *Patent Office Research and Development Reports* have indicated the importance of the effort accomplished.

The Patent Office studies can be divided among several groups. One of these (report no. 3) is an outline of a general character on the problems posed by the automatic retrieval of information concerning previous patents. Two reports (Nos. 6 and 10) contain an examination of the procedures to be followed with a certain type of machine (respectively the ILAS and the SEAC). Another report (no. 15) outlines a system of notation for the various characters, symbols, punctuation symbols, etc., encountered in patents, with a view to transcribing them on IBM machines 024 and 704. The foregoing is somewhat removed from our present subject.

A group of reports relate to 'coding in the field of chemistry', and more especially for research concerning steroids (nos. 2, 5,

7, 8, 11). The most interesting report, from our point of view, is report no. 13,¹³⁰ concerning recent studies undertaken on polymers, and which includes a general outline of 'logical' methods developed in conjunction with the ILAS punched-card machine and the Bendix G15D computer. This coding system contains a distinction between 'ingredients' (terms relating to chemical compounds), 'functions' (terminology which does not concern structure, and relating to processes, properties, reaction conditions, utilization), 'group' (component parts of a unit), and 'interrelations' (between the various groups).

The identification of chemical compounds is obtained by a method compatible with the so-called 'Variable Scope Search System' or 'VS3' described earlier in the report to the ICSI by the same three Patent Office researchers.¹³¹

The identification of functions—265 in number—is obtained by means of 'modulants', 2-character hexa-decimal prefixes which 'modify' the code word for the material.

The identification of groups is effected with 'grouping signals', which, in fact, serve as punctuation symbols.

Finally, the 'interfixes indicating the interrelations between the different groups'.

A very interesting procedure, which does not concern us directly here, however, is that of 'weighting' which consists in assigning a certain numerical value, called 'weight', to each group of subjects within a given series of questions, in order to arrive at not only an absolutely complete answer but the 'next best answer'. This method is directly related to searching strategy.

Still in the field of polymers, report no. 14¹³² describes another technique, used with the IBM RAMAC 305 machine, which is based on the principle of retrieval by 'characteristics' not by 'documents'.¹³³

The interesting aspect here is that three 'levels of descriptors' are used: the third or inferior for specific compounds, the second or intermediate for structural fragments of these compounds, the first or superior for the various 'common attributes of these fragments' (p. 5); one could moreover provide more than three levels (p. 6). We have here, therefore, a kind of hierarchical classification system, but the hierarchical relationships are not explained in the symbolization of the descriptive terms: should the machine encounter a generic term, it selects automatically the corresponding specific terms and thereafter conducts the search according to the latter (p. 5, also *Preprints ADIA*, p. 1).¹³⁴

A further series of Patent Office reports relates to 'linguistic problems encountered outside the chemical field'. They are primarily the result of extremely interesting research—from the standpoint of the present study—by Simon M. Newman.¹³⁵

Unfortunately, they cannot be summarized very easily. The three basic reports (nos. 1, 4, and 12)* are progress reports, corrected as the research progressed and giving the results of the latter, and no complete outline has yet been presented.

Newman undertakes the creation of a 'meta-language' which—according to S. C. Dodd—he has named 'Ruly English': a non-ambiguous language.¹³⁶

This meta-language contains six categories of elements—which we will arrange in a slightly different manner than that adopted by Newman himself (no. 12, p. 5):

1. *Roots and qualifiers or qualifying roots*
2. *Suffixes* referred to as *modulants*, added to the roots to define different aspects of the basic concept
3. *Quantification prefixes*, added to the qualifiers
4. *Interrelational concepts*, expressing the interaction between things or their mutual relationships, by means of a so-called 'distribution' operation. An interrelation concept is placed in apposition after each of the connected terms
5. *Numbers*, referred to as 'interfix' numbers, *suffixed* following the modulant or following the interrelationship concept: two or more than two descriptors (descriptor is Mooers' term, here it is used to mean a root + its modulant), which are in interrelationship or in interaction, are assigned the same arbitrary number, called an interfix, which signifies that they are connected; similarly an arbitrary and identical number—interfix—indicates that two or more than two concepts of interrelationship refer to the same relationship.
6. *Numbers*, referred to as *items*, identifying all descriptors relating to a same subject (which they describe in all its different aspects) and grouping them in articles of a numbered list; this numbering is called 'itemization', it serves more or less as an equivalent of the Perry-Kent code's punctuation symbols.

The roots, qualifiers, prefixes of quantification and concepts of interrelationship are derived from the English language by abbreviation, and at times by the combination of two abbreviations; the modulants consist in a 1- to 3-letter suffix (often mnemonic in English) and are separated from the root by the symbol =; the interfixes are added following the modulant or the interrelationship concept by a hyphen (-).

No extensive 'dictionary' of *roots* has been published; the list of those settled upon in 1958 is found, in alphabetical order (with that of the modulants, the interrelationship concepts and corresponding normal English words) in Appendix A of no. 12

* In all of the following references to these reports we shall indicate first the number of the report, followed by the page number or numbers.

(p. 8-14). Contrary to the method used in the Perry-Kent code, there is no breakdown into 'semantic factors'.¹³⁷

The *modulants* have been published in the form of a systematic table no. 4 (p. 10); but they have been revised in no. 12 (p. 6-7) and the revised list is that of no. 12 (p. 8); it is alphabetical; we prefer, here, to give the no. 4 (p. 10) systematic list, corrected, however, and explained according to the directions in no. 12 (the words in capital letters are the 'roots' of the 'Ruly English'):

- =NT *Process*: action or conduct which characterizes either (a) a property, an existing circumstance, an activity or mode of existence, or (b) a change in a property, an existing circumstance, an activity, or a mode of existence, of one or more substances, materials, or intangibles. A change in a property, etc., is considered as a change in its *state*
- =W *Work* (that which is worked)
- =WSM starting material
- =WIP intermediate product
- =WPF final product
- =WID ingredient descriptor
- =WCD component descriptor
- =WOR performer (device or object used in performing the work)
- =M Made from (source)
- =MSM source *substance*
- =MCC component of a complex
- =MCI combination comprising
- =SW Subcombination of whole
- =E *State* (meaning a property, existing circumstances or conditions, activities or the mode of existence which remain unchanged *during* a period of time, which may be indefinite or instantaneous)
 - =EP before the *process*
 - =ED *during* the *process*
 - =ER after the *process*
- =IS Illness of
- Numerical modulants:
 - =X or *more*
 - =Y exactly
 - =Z or *less*
 - =B used as an *ordinal* number

Explanations will be found in the original texts concerning certain of these 'modulants': notably on the meaning of 'Work' in the Patent Office's particular terminology, in no. 1, p. 5; concerning certain modulants in W and M, adapted following a proposition made by a Patent Office group working in the chemical field

(reproduced in no. 4, p. 16), in no. 4, p. 13; concerning 'Process', 'Device' and 'Work' in no. 12, p. 6-7. We have adapted as best we could the definition of 'State' given in no. 12, p. 13, but it did not seem to us particularly clear.

Newman states (in no. 4, p. 13) that the -IS modulant (illness of) 'is probably peculiar to medicine'; this seems rather curious, inasmuch as the notion of pathological state is in reality a notion which is extremely general in character.

In no place is found a complete special list of *interrelational concepts*. A list of cause/effect notions had been published in 1956 (tables 4 and 6, in no. 1, p. 6-7), but these terms are not found in the composite list of report no. 12. The terms relating to *temporal relationships*, of no. 1, p. 8 (table 8) have been re-used, substantially modified, in the list of report no. 12. Table 18 (no. 4, p. 6) relating to *relative measures* was modified in one respect only: SLI (little) and MAT (much, substantially) which were joined to 'less' and 'more' have become separate 'quantitative prefixes'.

Newman devoted a large part of his work on relationships to *spatial relationships*. He divided them into five categories: association, alignment, proximity, orientation, arrangement. Corresponding roots (with only minor changes) are found in tables 19-22 (no. 4, p. 7-9); for proximity, he uses the interrelational concepts of the category on 'relative measures'. Appendix A of no. 12 adds a few other relationships: comprising/consisting of, which is coded COMPRISOF (STOCKFROM) or CONSISTOF (STOCKFROM) according to whether the assembly contains the assembled parts in question as well as other parts, or those of the parts in question only (these two relationships, in a different form, were found in 4,11); a PENTHRU (THRUPEN) relationship which is that, for example, of a string *threading through* a rosary; it also revives a FROMWHENCE (WHENCE-FROM) relationship already encountered in no. 1, p. 9, concerning, for example, water poured from a pitcher.¹³⁸

From the foregoing examples it will have been observed that a similar relationship is designated by two different 'interrelational concepts', although the latter are connected by the parentheses symbol. Newman refers to these different names which express the two aspects of a same relationship, when the latter is dissymmetrical, as different 'mirror-image' interrelational concepts; it is not known exactly why he has not used the normal logical terminology (see no. 1, p. 4; no. 4, p. 5). He applies this same 'mirror-image descriptors' idea to oppositions in terms (no. 4, p. 16) which seem to have perplexed him; we shall not dwell on this aspect of the question (see Note 123).

Andrews and Newman (no. 1, p. 5-7) had undertaken (with C. G. Smith) an analysis of English prepositions (especially

'thru'—through, and other meanings); in no. 4, p. 5 they state that they have not continued in this direction and preferred a more direct approach. One finds also in no. 1, p. 8, an idea which does not seem to have been mentioned subsequently, concerning the fact (to be analysed) that 'the possibility exists that the codes for the modulants . . . are closely connected with codes for interrelational concepts'. In no. 1, p. 4-5, an analysis had been presented of the modulants of the 'ENPACKAGE' (packing, package) root which, it had been stated, had been previously defined in a non-ambiguous manner by the Patent Office's Classification Bulletin 402, in 1951 (this analysis can be compared with Foskett's classification mentioned above in Note 70), but, in no. 4, p. 14, it is stated that this concept finally revealed itself as implicating a whole series of different factors, overlapping each other, in correlation and non-exclusive; an analysis was therefore undertaken of specific modulants of less complex terms, concerning movement, change of form, assembly, substance, transfer of energy; only two rather brief pages have been devoted to these various fields.

In no. 16, Newman¹³⁹ resumed his analysis of prepositions, by studying those which begin in English with the letter *a*; he has retained 48 (counting as separate prepositions in this number 'about' and 'about.at', and similar cases, and including ten combinations with 'as') and he has distributed their different meanings under 31 headings called 'synthetic relational concepts' of the type already mentioned, AFTIM-TIMAFOR, etc. He has also regrouped them under the 44 headings of Roget's *Thesaurus of English Words and Phrases*. One may compare this work with that done on French prepositions by Ruvinschii (see Note 26).

As we have observed, the 'Newman language' does not make use of classification in the formation of its code-words, which appear as independent of each other (with the exception of those which have 'mirror images'). However, hierarchical relationships have been mentioned in several passages in the various reports (no. 1, p. 5; no. 9, p. 7-8; no. 12, p. 6). The fact that there are numerous different categories to which a notion can be related, is emphasized, with rather interesting examples. In no. 12, p. 14-15, is given a list of the various 'subsumed-inclusive relationships' which apply to the roots in Appendix A. These are 'weak hierarchies' in the Mooers sense. A mechanization of the process of establishing such lists is anticipated. Elsewhere (no. 9, p. 6-7), a discussion is found of the classification of 'manufacturing' methods in a broad sense, which could, states Newman, all be reduced to five mutually exclusive methods, and of the classification of apparatus on various bases (mainly the relative movement of the tool in relation to the object to be tooled, in so far as machines are concerned). Finally, under the name of 'compliance

coding' report no. 1 (p. 7) presented a coding method designed to indicate the degree of generality of a concept by columns, noting with 1 the presence of a property and with 0 (zero) its absence, these codes being organized in such fashion that the most specific concepts would have at least 1 in their code; the list of interrelational concepts in 4 has, in effect, a series of code-numbers of this type, numbering, according to the tables, from 3 to 10 binary figures, but these are no longer mentioned in report no. 12. Truthfully speaking, we do not quite see their utility—perhaps wrongly and through a lack of understanding.¹⁴⁰

It would certainly be premature to attempt even a broad judgement of Newman's work which, obviously, is still in process of evolution. It is a vast attempt at defining or re-defining concepts, which could perhaps be entitled—to paraphrase a famous title—'In search of lost simplicity': to discover or rediscover non-equivocal terms beyond the complications of natural language, which 'unfortunately' does not have 'uniform or logical rules for the denomination of devices or things' (no. 1, p. 3). Comparable to that of Proust, Newman's undertaking led him much farther than he thought in the beginning, and, no doubt, he has not yet reached the end of his efforts.

On the other hand, certain analogies with the Perry-Kent methods will have been noted. The 'modulants' bear some resemblance to the latter's 'role indicators', although report no. 1 (p. 12) mentions that this constitutes a mere meeting of thought and not a triggering influence, since the volume *Machine Literature Searching* had already been received at the time report no. 1 went to press. In the list given in report no. 4 (p. 10), a 'source' column, indicating comparisons or derivations, gives ten modulants adopted as identical to PK role indicators: six were to be found already in report no. 1, two had been suggested by the 'Chemical Task Force' of the Patent Office, and two have been borrowed. The fact that the modulants are suffixed and the role indicators are prefixed is of no great importance.

The 'Newman language' ranks at a lower 'degree of integration'¹⁴¹ of morphemes to that of the PK code, inasmuch as it does make use of the 'symbolization' of roots; no. 1 (p. 12) states that the distinction between analytic relationships and synthetic relationships was deemed unusable by the Patent Office: 'according to the latter's method of reasoning, an insect subjected to the action of an insecticide' (PK analytic symbol W) 'is both a "starting material" (synthetic code KAJ) and a "material subjected to a process" (synthetic code KEJ)'. However, as in the case of the PK code, this Patent Office language 'coalesces or merges',

according to the presence of modulants, pure inflexions attached to corresponding roots, and even 'interrelational concepts' which, fundamentally, have no existence of their own since they only appear in apposition to each of the terms which they serve to connect. Report no. 1, p. 3 is not too clear as to the reasons which led to the selection of this linguistic pattern.¹⁴²

It is true that linguists are not in agreement as to the relative superiority of the methods of expression of syntactic relationships. Although Frei or Sapir seem clearly to conclude in favour of 'pure syntactic relationship' systems; Walter von Wartburg entitles one of the chapters (most interesting, it must be said) of his work of synthesis 'the struggle for the maintenance of inflection in French'.¹⁴³ If one follows the school of this great Romance scholar, one may be inclined to consider that Perry and Kent, like Newman, were correct in adopting an inflexional system; as far as we are concerned—for reasons which cannot be developed at this point—we believe that the 'common language' of systems of information retrieval should be completely isolating, and we are therefore inclined to consider that the Leroy-Braffort 'linguistic pattern' is preferable to that of Newman (and, *a fortiori*, to that of Perry-Kent).¹⁴⁴

The name of Charles G. Smith has already been noted (p. 126); he collaborated in the analysis of prepositions in one of the first stages of the research undertaken by Newman. Smith was also a member of the Office of Research and Development, Patent Office, and also worked on a coding system designed for use with the ILAS machine, which he described in a very interesting report to the ICSI.¹⁴⁵

He has made use of arrowed diagrams, such as those of Leroy-Braffort (and also Jessica Melton, in the beginning, but the latter, after presenting her four diagrams for each of the four types of phrases, makes no further use of them). Beginning from a rather debatable view on relations between terms which, he writes, are not 'themselves terms' but 'a sort of abstract form in which different terms are placed',¹⁴⁶ he criticizes Andrews' and Newman's solution for expressing relationships, previously applied with ILAS,¹⁴⁷ 'of applying the relationship to each term to which it belongs and then connecting the terms to each other' which is tantamount to 'treating relationships as though they were a description of an object rather than a relationship between several objects'. He therefore outlined a method 'to express the relationship separately', as distinct from the connected terms, by a different utilization of interfixes which comes closer to the Leroy-Braffort method.

Smith subsequently outlines his ideas on the concepts to be used: his system, he states, seeks 'ultimate concepts . . . required in the definition of more specific concepts. . . . There is a basic layer of concepts which do not require definition. It is the use of such elemental concepts which is contemplated in the present system. . . . A fundamental feature [of the latter] is to seek beneath composite words the basic organization of elemental concepts which they represent, and to develop the essential combination for the definition of these words.'¹⁴⁸

The suggested system, according to Smith, 'does not identify the apparatus or the methods by applying functional denominations to them, but enumerates . . . the characteristics, such as a change from one state to another effected on the object in question, of which the interrelation necessarily characterizes them'; this amounts 'to describing the inventions rather than to classifying them according to their aspects'.¹⁴⁹ In order to illustrate this conception, Smith outlines his 'analysis of methods' ('all patents relate to a method'),¹⁵⁰ and outlines the manner in which he plans to describe machines, providing as an example a hydraulic auxiliary motor,¹⁵¹ which leads him to general considerations on the 'analysis of controls' and the coding of the latter.¹⁵² We can only refer the reader to the said report.

In his most recent report to the Conference at Western Reserve University, Cleveland, Don Andrews, in a discussion of the different methods of expressing relationships (but not those of Leroy-Braffort, Pagès, or Gardin) writes that this is merely 'the beginning of a long road' toward a satisfactory solution.¹⁵³ Recognition must be granted to the Office of Research and Development, which he directs, for attacking this difficult problem with determination, and for elaborating methods which, though inadequate as they may still seem, have definitely indicated the path.

We conclude this outline on the Patent Office by a reference to the very remarkable ideas outlined by Warren Weaver in his lecture at the American Patent Law Association on 15 April 1955¹⁵⁴ on the occasion of an examination of the 'Patent Office problems'. He indicated that, in his opinion, the *classification problem* was the central problem, and added that the complexity of a contemporary, large classification system such as the Patent Office classification could be only an artificial product, since this superficial complexity could be resolved into simple elements if the basic logical structure were considered in depth. Weaver gave geometry as an example, an apparently extremely complex structure, but of which the logical basis consists of a very small number of undefined terms and postulates: 'the entire remainder is merely the development of this simple core of an unbelievably

logical richness. . . . Would it be so very unreasonable to believe that, if someone were to examine deeply enough, with a sufficiently experienced and penetrating glance, the whole confused proliferation of facts and ideas which exists at the Patent Office, he could perceive their internal logical simplicity?'¹⁵⁵ The efforts of Andrews, Newman, Smith and others, indicate that the Office of Research and Development has courageously undertaken a search of this 'internal logical simplicity'. From this standpoint, the studies of the US Patent Office are certainly among the most remarkable we have encountered so far.

264 *IBM research programme. H. P. Luhn*

The IBM research programme is oriented rather toward the 'machine' aspect than toward the 'coding' aspect—as would be expected. However, a few references should be given here concerning work activities closer to our subject, and in particular those of Hans Peter Luhn.

The latter, aside from his studies on information retrieval machines, has devised various 'coding' methods, in the sense of special codes to be used as 'machine-language' for such or such a category of machines.¹⁵⁶ As we have already seen earlier (Note 113), he had advocated, as early as 1951, a method of characterization of a subject by a set of elements or criteria of identification, bearing a great similarity to those of Cordonnier and Pagès in France, and of Mooers in the United States. The 'process of broadening the concept' which he described was linked, in an interesting manner, to research strategy, and he proposed the compilation of a dictionary of notions where each of the specific terms not retained for indexing would appear under as many headings as necessary to represent it by the different 'key terms' where it would thus be indexed. Such a dictionary is very similar, in fact, to the *Thesaurus of English Words and Phrases* by Peter Mark Roget, published for the first time in London in 1852, and since that time it has been the custom to refer to the 'thesaurus' method.

This 'thesaurus' method renovated by Luhn has been applied, among others, to a coding experiment covering 1,200 technical reports undertaken by D. S. Tompkins at the IBM technical library, at Endicott, New York; the thesaurus used was subsequently adapted to another IBM technical library, that of the Military Products Division, at Owego, by C. Kuljian and D. Marr.¹⁵⁷

For about the last four years, Luhn has devoted his time to a large-scale research, aiming at no less than a *total* mechanization of the process of recording and retrieval of information.¹⁵⁸

This research has already led to promising results, perhaps more so in the field of indexing¹⁵⁹ than in the matter of automatic analysis or auto-abstracting.¹⁶⁰ A discussion of these subjects would exceed the bounds of the present report, but it should be noted here that they have (or may have in the near future) rather considerable implications, from the point of view of the present study—primarily—perhaps, because they would render it possible, eventually, to dispense entirely with any classification or any coding, thereby depriving our study of its object. But there is reason to believe that such a limit is probably inaccessible.¹⁶¹ More immediately, and also more certainly, the preparation of thesauri (which will remain for an indeterminate period an intellectual undertaking requiring the entire capacity of billions of neurons of human brains, which are also well constituted and well trained) will be considerably facilitated¹⁶² because the methods studied by Luhn and his team should anyway facilitate the coding work: notably by permitting the automatic compilation of lists of words and combinations of words (which is more important) with their relative frequency.

In another study, Luhn discussed methods of expression of relationships, and provides a method entirely different from that of the US Patent Office, usable on the IBM 101 machine.¹⁶³

We have already mentioned above (Note 134), J. J. Nolan's report concerning the use of the RAMAC 305; apart from the procedure we indicated to avoid the construction of pre-established hierarchies, replaced by groupings of *ad hoc* inclusion according to the questions, Nolan indicates (p. 9-11) further means of introducing 'groupings of phrases' (the equivalent of punctuation) and a translation in machine-language of the order of words.¹⁶⁴

Finally, we may mention here the very remarkable research conducted by T. T. Tanimoto in the field of what may be referred to as mechanical classification.¹⁶⁵

265 *Mortimer Taube and the 'Uniterms'*

Since 1952, Mortimer Taube has popularized, under the name 'Uniterms', an indexing method on cards 'by points of view', which in principle, is very similar to Mooers' descriptors, but which—at least in the beginning—did not show the same concern for precise definition of terms used, and did not use the Zator 'filtering technique'. With many users, Uniterms have tended to become mere 'catchwords' derived from titles or analyses, without even eliminating synonyms. An abundant literature is available, but we shall merely cite a few references on the subject, since the principle of this system is far removed from the preoccupations of

the present report. It should be noted, however, that among other users there is a tendency toward completing the alphabetical dictionary of Uniterms by lists of categories.¹⁶⁶

266 *Miscellaneous specialized coding systems: in fields other than chemistry*

We mentioned in our 1955 report (Unesco 320/5601, p. 25) the research programme on 'categorization' which is being conducted at Johns Hopkins University, in the field of *medicine*, by the Welch Medical Indexing Project; we have not been able to examine the unpublished reports on the subject, nor the 'Final report on machine methods for information selection' by W. A. Himwich, Eugene Garfield, H. G. Field, J. M. Whittok, and S. V. Larkey (1955); we therefore regret being unable to discuss them here at greater length.

In the supplement to our 1955 report to Unesco (Unesco 320/5X30) we alluded also to the research work undertaken by the Operations Research Committee of the American Bar Association, and, especially, to that performed by Vincent P. Biunno. We are indebted to him for a few documents which, in interest, relate directly to the present study.¹⁶⁷

Biunno foresaw the 'creation of a universal code' for the indexing of legal texts, which would have included a 'list of a relatively restricted number of basic terms, of which the combinations could express complex notions, whether of a generic or specific character' (memorandum dated 9 September 1954, p. 3).

In the field of *geology*, the code elaborated by W. I. Finch, at the US Geological Survey should be mentioned; this code is to be used in conjunction with pre-punched margin cards, which provide a categorization by types of rocks, types of deposits, etc.¹⁶⁸

We shall draw attention once more also to the categorization of the Office of Basic Instrumentation of the US Bureau of Standards in the field of *instrumentation*.¹⁶⁹

From the point of view of the present report, the most interesting document that we wish to mention in this paragraph relates to the code devised by Touloukian and his collaborators at the Thermophysical Properties Research Center¹⁷⁰ which contains essentially a *classification of substances* of every nature. Although extremely pragmatic and designed primarily to meet the particular needs of the centre, a study of it will be necessary if, as Ranganathan has been requesting for a long time, a category relating to various substances is to be elaborated for use in a series of classification schemes.

267 *Miscellaneous specialized coding systems: in the field of chemistry*

From the abundant series of coding systems in the field of chemistry, we shall merely mention a few typical cases.

A number of systems make use of classifications by separate categories, including those of the Bakelite Co.,¹⁷¹ the Merck Sharp and Dohme Co., Rahway,¹⁷² and the Socony Mobil Oil Co.¹⁷³

Others have applied 'role indicators', for instance the Linde Co.¹⁷⁴ and the Carbide and Carbon Chemicals Co.¹⁷⁵

Three systems call for a notation of chemical structures according to a topological principle.¹⁷⁶

The Opler method at the Dow Chemical Company has been adopted by the Midwest Research Institute, with a few modifications. Four magnetic tapes (IBM 704) are used for physical properties, uses, name and chemical structure of the chemical compounds, respectively, with a view to ascertaining the correlations between the physical and chemical characteristics of the latter with their known uses.¹⁷⁷

Opler has also outlined some ideas on 'syntactical diagrams' which are very similar to those later expressed by Leroy and Braffort, and on automatic translation of a natural language into a 'logical' language by means of these syntactical diagrams.¹⁷⁸

268 *Techniques aimed at using a more or less standardized form of natural language*

One of the present tendencies in the USA is to 'do without classification' and to study systems of information retrieval using the terms of natural language—or rather of one natural language, English—with the fewest possible modifications.

It is possible to see one of the origins of this state of mind in the long-established and marked preference of the Americans for cataloguing systems using alphabetical subject headings, but more recently there has been quite a different influence; that of research into mechanical translation.

The methods of the various teams or of isolated research workers, who might belong to this general group, vary greatly, as do their ambitions; besides, the inclusion of some of these teams in the present group would be somewhat doubtful, and depends upon what one understands by 'natural language'.

In the provisional French version of the present work we included the name of Mr. Taube amongst those mentioned in this chapter; Dr. Burton Adkinson objected that the Uniterms method involves the selection of terms used for indexing by men (and not by machines), and that they are isolated terms, 'therefore', says Dr. Adkinson, 'isolated words do not constitute a language'.

Certainly! However, this Uniterms method uses combinations of words drawn (without codification or classification) from the natural language of documents (or from summaries or translations of them), and these combinations of words are nearer to natural language than, shall we say, the numbers of the decimal classification or the 'semantic factors' of Perry and Kent. It must also be remarked that, even if the key words—let us refer thus to the Uniterms—can in principle be combined freely with each other, in fact these combinations are limited, and the 'combinatory power' of the different key words is very variable. Claire Schultz and Clayton A. Shepherd have studied, using a Univac calculating machine, this combining power of the key words used in the document research system of the Merck Sharp and Dohme Corporation,¹⁷⁹ and, by a similar method, a sample of about 20 per cent of the material of ASTIA.¹⁸⁰ Lauren Doyle and John Olney have, for their part, made a study of 'interrelationships in word frequencies in topically close articles' on an IBM 709: one is here nearer to natural language, for it is no longer a matter of key words already selected by a documentalist, but rather of the words of the document itself.¹⁸¹

One passes, by various transitions, from methods using isolated words, or words joined in twos or threes, etc., just as they are found in the texts, or selected by the intervention of a human operator, to those which consist, no longer of key words, but of key phrases, forming a more or less standardized language, always based, however, on natural language. The first attempt of this kind was no doubt that of Krieger in 1949;¹⁸² one finds since then various attempts in this direction, of which the first seems to have been that of Miss Phyllis M. Williams who worked at first independently and in liaison with ACF Electronics (which later became Avion Division), and then joined the team of the Itek Corporation in 1958.¹⁸³

It is difficult, without having sufficient knowledge, for us to describe in detail the work of Miss Williams, or that of R. P. Mitchell at the Lockheed Aircraft Corporation (Missiles and Space Division, Sunnyvale, California), who studied there a 'standardized English' under contract with the Rome Air Development Center.¹⁸⁴

The research at the University of Pennsylvania, directed by Zellig S. Harris, leader of one of the most influential (and most controversial) of the American schools of structural linguistics, is based on his theory of 'transformations', first explained in an article of 1952,¹⁸⁵ and then set out in detail in 1957.¹⁸⁶ In the field which interests us here this has been subsidized by the National Science Foundation (and to a considerable extent), since October 1956.¹⁸⁷

The most accessible report on the work of the University of Pennsylvania team is the paper read by Harris to the International Conference on Scientific Information,¹⁸⁸ but the details must be looked for in internal reports, of which twenty-seven have already been published.¹⁸⁹ The general results have been formulated as follows:¹⁹⁰ (a) experimental analysis of the size of kernels, or subsentences, most appropriate for information retrieval; (b) experimental analysis of the classes and subclasses of English words suitable for making a syntactical analysis which can be used on machines; (c) compilation of a grammar of the English language formulated with the aid of recursive operations; and (d) a method for recognizing the syntactical structure of any English phrase, by means of a calculating machine. This method has led to the compilation of a code, used at the present time with a Univac calculating machine. Other research is being done:

(a) generalization of the mathematical theory, of the linguistic methods, of the programming methods and of the existing possibilities of automatic coding, by starting from the syntactical analysis already set out, but generalizing it for any language or pseudo-linguistic system and for any type of calculating machine; (b) a programme for carrying out an analysis of transformations, based on the present programme of constituent-analysis, but permitting the cutting down of phrases into subphrases, and transformations; and (c) a more complete description of the subclasses of words and of transformations in English, and the preparation of a transformational word-list (dictionary) for English.

One sees that, contrary to the research mentioned at the beginning of this section, Harris (like Mitchell), concentrated in the first place, and even exclusively, on the study of syntax. Noam Chomsky, who was a pupil of Harris, worked out and published a more complete, and somewhat different theory; the book has already become a classic.¹⁹¹ There seems to be a fairly close connexion between the purely theoretical work of Chomsky and that of Victor H. Yngve, which is directed more towards applications. The latter has concerned himself mainly with mechanical translation, but, with the support of the National Bureau of Standards, has conducted research 'on the possibility of using natural language for the storage and retrieval of information in a mechanized documentation system'.¹⁹² The two reports published up till now are of a too general character to permit one to judge this work.

The reports by Harris and by Chomsky have attracted the attention of French¹⁹³ and Russian¹⁹⁴ research workers.

Another method for the 'determination of the structure of

phrases' which may be mentioned here—although it is aimed at applications to mechanical translation—is that of the Rand Corporation. Hays¹⁹⁵ shows clearly that this method can be applied to the 'automatic determination of categories for indexing' of documents.

The group at the Smith Kline and French Laboratories, Philadelphia, has worked out a system of indexing documents including an 'arbitrary and synthetic grammar'; its opinion is that 'the syntax of natural languages is too complex and too variable to be used as a basic grammar in automatic indexing'.¹⁹⁶

The Data Processing Systems Division of the Research Information Center and Advisory Service on Information Processing, at the National Bureau of Standards, directed by S. N. Alexander, was inspired by the research of Chomsky, Harris and others, and considers that 'the solutions found to syntactical problems may bring great progress in the search for a solution to the problem of automatic documentation'¹⁹⁷ in fact several National Bureau of Standards reports deal with the field of the theory of syntactical types.¹⁹⁸

Ramo-Wooldridge is running a study, subsidized by the Rome Air Development Center and the Council on Library Resources, with the eventual object of machine-made abstracts of scientific texts; the first phase of this research was recorded in a report of a rather general character.¹⁹⁹

The research in question consisted in examining the degree of relevance of 100 articles in relation to 50 questions, this degree being judged by physicists using a 'weighting factor'. The articles were catalogued by a standard method of alphabetical subject headings as a control; in addition two methods for automatic retrieval from documents were worked out—one method involved simply the selection of documents containing certain combinations words or of groups of words corresponding to the various questions asked (e.g., all those containing the phrase 'charge polarization', or 'charge distribution' at the same time as the words 'scattering', 'scattered', or 'scatter'), and the other method, also included recourse to a thesaurus, which permitted the person asking the question to formulate it in a different way.²⁰⁰

Swanson's experiment included no attempt to determine syntactical relations; in fact, he suggests substituting for syntax indications of 'proximity' (i.e., of physical distance in the text) of the words indicated.²⁰¹

None of the methods used has produced satisfactory results, mechanical searching, however, being more efficient than the search method not using it.

We shall finish by quoting just two more items of research, those of F. W. Householder²⁰² and of Ron Manly²⁰³, who ought,

it seems, to join the 'family' presented in this section, but on whom we have no information other than the very short accounts given in the last number of *Current research and development in scientific documentation*.

As can be seen, the tendency which we have tried very briefly to characterize here in its main outlines seems to be to recruit experts rapidly, and this not only in the USA (where it seems to have had its origin), but also in Europe. It has at times been in quite distinct opposition to all attempts to elaborate systems of information retrieval using systematic codes,²⁰⁴ and if the advocates of this extreme view were right, the very object of the present work could be considered as belonging to the past. One may conclude also that, even if 'artificially' simplified languages, including a minimum of syntax, and using systematic classifications of concepts, and/or the reduction of the simple concepts to elementary components, can serve good purposes today, the future is nevertheless with the 'completely automatic documentation' methods, which tackle directly texts written in any 'natural' language, however complex this language may be. This is approximately Leroy's attitude, in which he is more or less opposed to that of Gardin.²⁰⁵ Bar-Hillel, on the contrary, has reached the conclusion that completely automatic documentation is impossible.²⁰⁶ We should not like to be as positive as him,²⁰⁷ but we must admit a certain scepticism towards the projects for the mechanization of information retrieval based essentially, and almost exclusively, on the analysis of the syntactical construction of 'natural' languages.²⁰⁸

269 *Two special studies*

Under this somewhat enigmatic title, we plan to dwell briefly on two studies which are extremely difficult to classify.

The first is that made by Miss Mary E. Stevens, at the National Bureau of Standards. The so-called 'IQ' research consists in exploring the possibilities of the SEAC machine to perform operations of logical recall of recorded information, the latter consisting of words (either in normal English, or coded) accompanied by an index of indications relating to semantic and logical interrelationships between the terms selected. The machine can therefore: 'define', by supplying the generic and descriptive terms related to the term of which the definition is sought; 'develop', by furnishing specific examples of a generic term; 'localize', by indicating the place which can be associated with the proposed concept; 'match', by comparing several proposed terms together, in order to find a 'common point' making it possible to relate to these terms another term possessing the same characteristic; and

carry out other logical operations (reject, select, learn, etc.). These experiments, of course, are still rudimentary—but they doubtless lead the way to a future of great promise, and of which Mary Stevens indicates a few of the possibilities in the field of document retrieval.²⁰⁹

The second research being undertaken along new lines is that of M. M. Kessler at the Lincoln Laboratory, Massachusetts Institute of Technology, based on the hypothesis that an indexing system could be established independently of words and language (normal): from 'scales of magnitude concerned in the document', from its position in relation to an 'experimental-theoretical axis', etc. The brief report at present available gives a few details only on the method followed, but the basic idea is most interesting and likely to lead to important developments.²¹⁰

27 Soviet Union

Unfortunately, the documentation which is available concerning research in the USSR in the field covered by the present report is very fragmentary, and we must stress the fact that without any doubt a more complete study should be made later.

In fact, the only Russian research concerning which we are well informed is that of V. P. Čerenin and his collaborators at the Computing Center of the USSR Academy of Sciences.²¹¹

Čerenin's research began in 1954, both on 'informational language' (Nekotorye problemy . . .) and on a so-called EIM punched-card machine, as well as a code (in the sense of a machine-language code) for use with such a machine.²¹² In 1958, the machine was modified in order to meet the increasing needs created by the development of the contemplated coding method (in the broad meaning of 'informational language', which is Čerenin's term for it). The latter was originally inspired by the first Perry-Kent studies (articles in *American Documentation*, 1954) and later, influenced by Andrews-Newman, advanced toward a more complex form than that envisaged in the beginning.²¹³

These modifications in EIM (see *EIL*,* p. 413-5, made possible in 1958 the first experimental retrieval of documents in the field of mechanics.

Approximately 1,000 abstracts gathered from the abstract bulletin of the Soviet Union's Academy of Sciences, mechanics section, for 1953, were indexed. A preliminary indexing (*EIL*, p. 415-6) provided approximately 1,500 titles of subjects, which were

* *EIL* is the abbreviation for the report by Čerenin, Lavrent'eva and Zidkova, 'An experimental informational language . . .'. (See Note 211.)

analysed in individual terms, ending in a list of approximately 4,000 terms (including synonyms).²¹⁴ These terms were divided into *five general categories*:

Processes and states: movement and equilibrium of bodies and apparatus, efforts and distortions, regulatory and control processes, etc.

Objects: bodies (solids, liquids, gases, mixtures), various devices (mechanisms, machines, instruments, tools, accessories) and installations; the solid bodies class includes various materials (metals, wood, cement, etc.)

Properties (of bodies and materials, of the various devices and processes, etc.) and sizes

Abstract concepts and other terms, not included in the preceding categories

Proper names.²¹⁵

A system of definitions and of cross references from the synonyms to the terms retained was subsequently established (*EIL*, p. 417-8). On the basis of its definition, to each of the special terms was then assigned a certain number of selected characteristics, previously identified, linked analytically to this term. For instance, paper is analysed into solid, paper; aluminum into solid, metal, aluminum; rotation into movement, rotation; thermometer into device, measurement, temperature. Certain concepts which can be linked to various categories form the subject of various 'representations', each of which is employed according to the point of view to be expressed, e.g., sand will be linked, according to the problems studied, to springy foundations, but also to porous materials and crumbly bodies. These 'developed formulae'—as we may name these (this term is not contained in the text of the Russian authors)—not only contain the generic characteristics of the basic term, but also the characteristics which have with the latter an association of 'similarity'. Zidkova, provides, among other examples (*EIL*, p. 419) shock-absorber: device, elimination, vibrations. A system of cross references by 'see also . . .' completes the indexing method which, fundamentally, bears a strong resemblance at this stage to the older classical catalogues of the 'alphabetico-systematic' type; it will be noted that the analysis and the classifications which are included are effected from the special standpoint of mechanics (see the example for paper, *EIL*, p. 418). This language is a definitely specialized language, much narrower than that of Perry-Kent. All of the above constitutes what the authors refer to as 'primary indexing'.

The 'secondary indexing' divides into two categories—for the purpose of transcribing onto punched cards—the terms retained during the primary indexing of an article: 'basic subjects' (i.e. objects and processes) on the right-hand side of the

card, and on the left-hand side, attributes, methods and operations. The characteristic terms are coded (in the narrow sense of the word) in code words of four letters of the cyrillic alphabet selected preferably for their mnemonic value (in Russian, of course): with the exception that here there is no search for autodemarcative words, this amounts to the Perry-Kent system (*EIL*, p. 406, p. 412-13 and p. 422). In the case of sizes, the first three letters of the code word indicate the size, and the fourth its value according to a conventional scale (e.g., *z* = zero; *l* = large; *s* = small). Great speed for instance would be *spel*, reduced speed, *spes* (*EIL*, p. 410 and p. 422).

What would appear to be the most important feature of the *EIL* method, is that the 'developed terms' (in associated characteristics) which result from the analysis are connected into groups by a set of parentheses, according to a 'structural formula' expressing the content of the article. The 'permanent core' of this formula is a 'phenomenon', although it may 'degenerate into a simple indication of a process or of an object'; in general, 'the phenomenon can describe the interaction of several objects and can include the various processes specifying this interaction'.²¹⁶

This set of grouping parentheses is explained in regard to the articles dealing with movement: the first parentheses contain the word 'movement' and its particular characteristics; the last parentheses relate to the moving object, also characterized; the intermediate parentheses contain the 'other objects', with their characteristics (*EIL*, p. 419-20).

The 'active' object is indicated by the symbol \times and the 'passive' object by $\times\text{x}$. (*EIL*, p. 423-4.)

In research strategy, questions should be put by using the same rules of order; the degree of generality of questions may, however, vary (*EIL*, p. 424-5).

Čerenin avers (*EIL*, p. 401-2) that his 'experimental language' is simpler than the Perry-Kent and Andrews-Newman languages, the method of expression of relationships by simple groupings being less complex. He concludes by mentioning certain difficulties encountered in the research, which would probably motivate the subdivision of the movement characteristic into 'movement in', 'movement outside of', etc. Other modifications could be effected in the system, which is still in the experimental stage (*EIL*, p. 426-7).

There is no doubt that we have here an ingenious and economical solution, whose developments should be followed with the greatest attention.

From the A. M. Zuckermann and A. P. Terentiev report

to the Cleveland Conference,²¹⁷ it would appear that the question of the coding of chemical structures has been the subject of extensive research in the USSR, by the authors of the report themselves and by others, particularly G. Vleduc. We shall not follow this subject further since it is more specialized than the present report.

All that remains is to make a brief reference to the research being performed by L. I. Gutenmakher which seems, however, to be directed rather toward the information machines aspect than that of codification.²¹⁸

Through lack of sufficient documentation, we have been unable to include in this study Eastern countries other than the Soviet Union; we are the first to deplore this gap.²¹⁹

28 India

We shall refer briefly to the work performed by the *Library Research Circle* led by S. R. Ranganathan at the University of Delhi Library and the *Insdoc* subcommittee on classification at New Delhi. The research conducted by this group is very closely related to that performed by Ranganathan, and constitutes for the present, in a general sense, extensions of the Colon Classification in specialized fields based on the same principles: agriculture, medicine, labour economics, demography. Special reference should be made to the studies of a group, of which B. V. R. Rao is the secretary, constituted in July 1958 and engaged in the preparation of a complete list of 'properties, values, and operations'.²²⁰

29 Japan

At the Cleveland Conference in 1959 Kinzô Tanabe presented a report on an experiment in mechanical selection of documents in the field of diesel engines, performed by a team of three researchers (assisted by seven students) at the Japan Information Center of Science and Technology.²²¹ The code employed was of a type similar to those of the English Classification Research Group, with eleven 'facets'; it is provided in an appendix to this report. There has been no research on the expression of relationships.

3 General categories and the expression of relationships in natural languages, and experiments in international auxiliary languages

As indicated in the Introduction (p. 13 and 15-6) a very limited outline only can be provided here—amounting to scarcely more than a mere preliminary bibliography—a provisional draft of a more important work, which we hope to publish later.

31 Studies on categories and relationships in natural languages

As we have already pointed out several times above (and notably in section 268) the importance of the study of 'natural' languages in relation to information retrieval is being more and more clearly recognized today.

This movement has been stimulated by the research being conducted on automatic translation; to a certain extent it constitutes a development or a branch of the latter. However, as Andreev has so well demonstrated¹ automatic information retrieval (or automatic documentation) presents, in relation to automatic translation, some characteristics which have not always been sufficiently noted—making it a more difficult task in certain ways.² For, in this case, 'not only does the formal structure of the message cease to be invariable, but its content too'.

There is no doubt that, if rational methods of automatic information retrieval are to be established, a study of the linguistic structure of documents from which the information is to be drawn cannot be neglected—regardless of the particular method used. Furthermore, one of the problems which documentation must certainly solve in the future (and perhaps its most important problem) would seem to be that of devising a rational ‘documentary language’ standardized on an international scale. Hypothetically, one can imagine that such a language must be inspired as much by the processes of expression of natural languages (or at least by some of these processes) as by classification or coding systems worked out in the past independently of lessons drawn from general research in linguistics.

The above statements justify the presence, in this study, of a chapter devoted to ‘categories and relationships in natural languages’. *A priori* one might conclude (and that was our feeling in the beginning) that the drafting of this chapter would be facilitated by the considerable development in linguistic studies during the past nearly hundred and fifty years since the existence of a comparative grammar accepted as scientific (if the 1814 Rasmus Rask paper is acknowledged as the date of origin of the latter; but perhaps it would be more accurate, here again, to go back as far as 1710, to Leibniz’s *Brevis designatio meditationum de originibus gentium ductis potissimum ex indicio linguarum*). It must be recognized, however, that, unfortunately, we are far from having at our disposal a general treatise on the subject: those available are hardly more than general outlines of a predominantly theoretical character, or manuals of materials in a more or less ‘rough’ (or semi-finished) state which provide either descriptions of languages or more elaborate studies concentrating on questions of detail.³

It should also be observed that contemporary linguistics is itself a sort of ‘Tower of Babel’, with various opposing theories which do not even use a unified terminology. This situation justifies the nature of the present preliminary report, which consists in pointing out a few of the directions in which research is at present being conducted in this field, but it should be considered merely as an orientation; we are only too aware of its gaps and insufficient precision.⁴

In view of the habitual divisions between more or less opposing or rival schools of linguists, and of which the frontiers correspond rather closely—in general—to their country of origin, our outline will be set out primarily in geographical form.⁵

311 *Germany*

Historically, it is in Germany, in the works of Wilhelm von

Humboldt, that we find the first important studies on the structural form of language, and his linguistic work could well be re-examined as a source of inspiration.⁶

However, Bopp and his successors were to lead German linguistics into another direction and dedicate it for a whole century to purely historical studies; this stage was probably a necessary one, during which results were achieved which should not now be called into question, but which should be surpassed.⁷

A. Marty⁸ and Wilhelm Wundt⁹ were to appear on the scene before other avenues could be explored more or less successfully, and two more decades were to elapse before Cassirer's research¹⁰ followed by the appearance of the theory of 'semantic fields';¹¹ but all this, however, remains outside the scope of a study devoted purely to grammatical structure.¹²

312 *France and Switzerland*

The 'Franco-Swiss school' is frequently referred to, by reason of the close contacts which existed between Ferdinand de Saussure and the Parisian linguists, before the man considered as their master by most modern 'structuralists' returned to settle in his native Geneva. There is, moreover, an almost perfect coincidence between the date when Saussure began, in 1906, to make known to his Geneva students—his ideas on general linguistics and in particular, his ideas on 'general synchronic linguistics'—and that of Meillet's opening lesson, succeeding Bréal at the Collège de France, on 13 February 1906, where we find the famous phrase 'l'histoire ne saurait être pour la linguistique qu'un moyen, non une fin' [for linguistics, history should be a means, not an end].¹³

In the perspective of time, however, it may seem that Saussure was the innovator: Meillet would seem rather to be the last of the great 'comparativists', and Saussure the first of the great structuralists.¹⁴

It still holds that very few linguistic texts have had a more lasting influence than the fifty or so pages in which Saussure outlined the principles of his 'synchronic linguistics'¹⁵ they are so well known, and their substance has passed so well into the 'common fund' of the present science of language, that it is not necessary to go into further detail here. The basic idea of oppositions (*Cours de Linguistique Générale*, p. 149, p. 164 and p. 168), the complementary notions of relationship (*ibid.*, p. 170-5), of solidarity and groupings (*ibid.*, p. 176-80), the diagram of the mechanism of the 'limitation of the arbitrary' in language (*ibid.*, p. 180-4) or that of the abstraction of grammatical notions (*ibid.*, p. 189-90)—all of the foregoing has now become independent of

its Saussurian origin and has acquired a general scientific value.¹⁶

In Switzerland, the research of Saussure has been continued by Charles Bally¹⁷ and A. Séchehaye¹⁸ and subsequently by H. Frei.¹⁹

Returning to France, the kind of interlude created by Ferdinand Brunot in his attempt to analyse a language (French) on the basis of expressed thought ('rational concepts') should be mentioned. But, however interesting the analyses of detail may be, we are here far away from contemporary structuralism.²⁰

Chronologically, Gustave Guillaume could have taken rank slightly before Brunot, inasmuch as his first great study was dated 1919, but his actual influence is much more recent—and continues, since there is a 'Guillaumian' school (and Guillaume himself speaks of 'Guillaumian linguistics'). Such a 'sub-linguistic scheme' may be severely looked upon²¹ or considered a rather disconcerting terminology.²² Even more disquieting, probably, are the metaphysical concepts which are expressed by certain general formulae of the author.²³ Consequently, a detailed critique (which would not be appropriate in this report) would reveal in certain 'Guillaumian' analyses—in particular those relating to temporal categories—some rather ingenious theoretical views, which would, however, still have to be subjected to a verification test regarding other languages than those (substantially French, plus a little Greek and Latin) which were used by Guillaume in building up his theories.²⁴

In 1934 there began to emerge, in the lectures at the Institute of Linguistics of the University of Paris, a tendency toward describing the 'grammatical system', or the 'structure', of a certain number of languages, and a series of monographs were published. They were inaugurated by A. Basset on the Berber language, and Meillet himself took part (with E. Pichon: 'Structure générale du français d'aujourd'hui' 1935); two were published in 1937, one in 1949, two in 1950-51, and three in 1952-53. Georges Gougenheim published in 1938 his *Système grammatical de la langue française*, Paris, D'Artrey, in which he refers to the works on which Trubeckoj had founded phonology on the study of distinctive oppositions and wrote that he had 'applied a similar method to the study of grammatical elements' (p. 9). Similarly Gérald Antoine, in his thesis (see Note 20), attempts to 'express' a similar technique in 'a subdivision of syntax'. Somewhat isolated in Montpellier, Lucien Tesnière worked for twenty years towards improving a method of representation of syntactic relationships by 'stemmas' which he had formulated in 1932; this effort resulted in a considerable posthumous work.²⁵ Emile Benveniste had been, at the Congress of Linguists in Copenhagen in 1936, one of three participants in a discussion of Hjelmslev's paper (which, in the

perspective of time, can certainly be qualified as historical, but which—according to M. Cohen's report to me—had surprised rather than created interest); he became, in 1949, one of the two French collaborators in the 'Structural research' offered to the Danish master on the occasion of his fiftieth anniversary.²⁶ Finally the *École Pratique des Hautes Études* recently established for André Martinet the first French chair in structural linguistics.²⁷

313 *Denmark*

Copenhagen is certainly one of the main centres of contemporary structural linguistics: since 1939 the sole journal is published there whose title (*Acta linguistica, international review of structural linguistics*) specifically mentions the notion of structure. According to Hjelmslev, it was perhaps here that it was actually born, in the work of H. G. Wiwel.²⁸ Furthermore, Otto Jespersen (1860-1943) can reasonably be considered as being one of the originators of one of the master concepts of Hjelmslevian glossematics, the fundamental importance assigned to the notion of direction.²⁹

Viggo Brøndal's work³⁰ can be interpolated here; he doubtless owes much to Jespersen and to his ideas on the 'logic of grammar', but had been greatly influenced by the Prague school (Trubeckoj, Jakobson) and its theory of oppositions. Brøndal himself indicated very clearly the bases of his doctrine, which 'consist in recapturing in language the concepts of logic'. He discovered two categories of concepts: rational concepts (symmetry, transitivity, connectivity, variability, plurality, generality, continuity, totality, extension, integrity, universality) and generic concepts (relationship and object, quality and quantity).³¹

This attempt has been severely criticized,³² and it must be recognized that his constructions often convey the impression of being the product of an *a priori* theory to which facts are shaped for better or for worse.³³

It does not seem necessary to discuss Louis Hjelmslev's glossematics here at any length. It is doubtless the most consequential theory, the most precise in its terminology, for the formal description of language. It cannot be summarized into a few simple propositions, precisely because of its strict exactitude. However, there is no systematic and detailed outline of the latter, which would constitute a convenient 'manual' and would present a sufficient number of examples of applications; this is an urgently needed undertaking.³⁴ Meanwhile, one can only advise a reading of Hjelmslev's publications in their entirety (which, moreover, clarify each other) as well as of the applications of his method which have been made to the description of various states of languages.³⁵

314 *Great Britain*

The Irishman James Byrne published in London in 1885 a book entitled *General principles of the structure of language*, which was much appreciated by Whorf but criticized by Brøndal;³⁶ we have not seen it. And, truthfully speaking, we admit our lack of documentation on the present trends of English linguistics, apart from the research being conducted at Cambridge by the Language Research Unit, which will be mentioned later with regard to questions of semantics and lexicology. It seems, however, that the structuralist movement has not penetrated very deeply into Britain.³⁷

315 *United States*

On the subject of American linguistics, there is a report prepared in 1950 by John B. Carroll for the Carnegie Corporation, later revised in 1951-52 and published in book form,³⁸ containing quite a number of references; a more recent manual by Charles F. Hockett provides an 'average' comprehensive view, for the use of students, and is accompanied by a bibliography.³⁹

The 'luck' of the American linguists has been to find, on the territory of their own nation and that of neighbouring countries of easy access (Mexico, etc.) a variety of indigenous linguistic structures, greater than almost anywhere else. The necessity of describing very diverse languages, often far removed from the familiar Indo-European or Hamito-Semitic types, led them to seek more general and more objective descriptive methods. It is therefore not by chance that modern linguistics begins in the United States with Franz Boas' great handbook.⁴⁰

The half-century which has elapsed since the publication of this book has been dominated by the influence of Leonard Bloomfield⁴¹ and of Edward Sapir.⁴²

The writers who come closest to Edward Sapir are Harry Hoijer,⁴³ Benjamin Lee Whorf,⁴⁴ and Joseph H. Greenberg;⁴⁵ we shall discuss this further when examining briefly the problem of the relationships between language and culture (p. 152-3). Following Bloomfield's teaching one finds: Zellig S. Harris⁴⁶ and, through the latter, Charles C. Fries⁴⁷ and even Noam Chomsky⁴⁸.

We can include the name of Roman Jakobson in this section, since he has resided in the United States for many years, after having been with N. Trubeckoj, who was before 1939 one of the more brilliant members of the Linguistic Circle of Prague. Apart from his research in phonology, which has no relevance to the present study, he applied the method of distinctive oppositions

to morphology in certain important works which have created a school.⁴⁹

Finally, a certain 'humanistic reaction' (as it may be described) has been noted in the United States against the abuses of too rigid a 'mechanism'. It expresses itself clearly in Joshua Whatmough's book,⁵⁰ which is the most recent outline of synthesis on language, and also in an article by Erwin Reifler.⁵¹

316 *Italy*

It does not seem that Italian linguistic circles have shown much interest in synchronic studies (to use a Saussurian term): they include eminent researchers in historical linguistics, in etymology, etc., but so far the contemporary structuralist movement does not seem to have exercised a notable influence upon them.

Nevertheless, mention should be made of some works originating not from a linguist, but from a philosopher, Silvio Ceccato, Director of the Centre of Cybernetics and Linguistic Activities at the (State) University of Milan,⁵² which have been applied to research on mechanical translation⁵³ and have been the subject of preliminary experiments, unfortunately not continued, through lack of time, on the automatic making of summaries.⁵⁴

317 *Poland, Hungary, Czechoslovakia, Rumania, USSR, Yugoslavia, China*

In Poland, J. Kurylowicz has been one of the first partisans of glossematics,⁵⁵ and in 1955 he gave a general outline of the state of research in general linguistics in his country.⁵⁶

We have not seen two reports by the Hungarians J. Herman and I. Papp, of which it should be interesting to have a translation.⁵⁷ The book by M. Gábor (*Nemzetközi fordítókulcs*, Budapest, 1957) should be noted; its interest was emphasized by I. A. Mel'čuk⁵⁸ in a review.

In Czechoslovakia the 'School of Prague' was, before the war, with Copenhagen the principal centre of structuralism.^{58a} It is continued today by B. Trnka⁵⁹ and one may observe that it has given birth there to a group of young researchers who dedicate their work to mechanical translation and to applied linguistics, under the impetus of Petr Sgall and of Bohumil Palek.⁶⁰

In Rumania, L. Sainénu (who signed himself later as L. Sainéan in studies published in French, notably on slang) had published as early as 1891 a study on the relationships between grammar and logic;⁶¹ this study is cited several times in Hjelmslev's *Principes de grammaire générale*; we have been unable to

examine it, and we must also admit our ignorance of present-day Rumanian research, having merely heard the names of Professors Graur at Bucharest and Petrovici at Cluj.

In the above-mentioned book (p. 111), Hjelmslev speaks highly of the 'Russian school' which, following in the steps of Fortunatov, has applied strict methods to the study of Russian grammar, and he cites in the bibliography works published between 1881 and 1923 by Buslaev, Peškovskij, Peterson, Poržen-sin'ski and Potebnja. As is already known, Soviet general linguistics was for some years officially dominated by the influence of N. J. Marr, until such time as the latter's theories, preoccupied rather in establishing the monogenesis of language, were condemned in 1950 by J. Stalin.⁶² Since then, there has been a revival and considerable discussion has taken place, in particular following the creation, in 1952, of the *Voprosy Jazykoznanija* [Problems of linguistics]. Various trends have manifested themselves: e.g., a study by V. Z. Panfilov (a philosopher and not a linguist) shows a profound mistrust of structuralist endeavours, which seems to be based on a profound ignorance of their contents,⁶³ but the questionnaire 'For a discussion of problems of structuralism' published by the *Voprosy Jazykoznanija* has proved the desire of its writers to seriously explore all of the aspects of the question.⁶⁴

In the last few years this intense revival movement has shown itself in connexion with the development of research on mechanical translation and 'information languages'.⁶⁵ We quote in particular the names of I. and I. Revzin,⁶⁶ I. A. Mel'-čuk,⁶⁷ N. D. Andreev,⁶⁸ O. S. Kulagina,⁶⁹ and S. K. Šaumjan.⁷⁰ After overcoming fairly violent opposition from numerous linguists—who condemned (and no doubt still do) the structuralist methods as being connected with 'bourgeois' philosophy—this movement seems truly now to have won a position which is without doubt still destined to be enlarged.⁷¹

Since we have available only the note by N. D. Andreev in *Voprosy Jazykoznanija*, no. 4, 1960, p. 165 (English translation JPRS 6732, p. 46-7), we can say very little here about the research being done in Yugoslavia, except that it seems to be active, and stimulated by the work on mechanical translation (cf. an article by D. Popovich in *Borba*, 14 August 1960 and the translation JPRS 6732, p. 48-50).

As far as China is concerned, we regret that we possess insufficient information; it consists mainly in the report presented by Lo Tchang-Pei and Liu Chou-Sang to the International Conference for the Unification of the Language, October 1956 (translated by André Lévy in *Recherches Internationales*, no. 7, May-June 1958, p. 94-126). It is a very interesting document on questions of linguistic standardization, demonstrating a practical

and realistic viewpoint, as well as a great concern for the establishment of a standardized grammar based on scientific study.⁷²

In 1958, structural linguistics was the subject of an article, of which we have seen only an abstract;⁷³ there is no doubt that, in China as elsewhere, the research on mechanical translation begun in 1958 will give a great impetus to the development of structuralist methods.⁷⁴

318 *Some types of linguistic categories*

In the present chapter, our primary idea had been to examine in detail the operation of certain linguistic categories; we recognized very rapidly that such an ambition by far exceeded both the limits of the present report and the possibilities offered by available documentation on the subject, which includes an immense mass of detailed research and very few syntheses (and, it should be added, even fewer syntheses based on a sufficiently exhaustive statement of facts). Furthermore, it seemed necessary first to characterize various tendencies in contemporary linguistics, and this is what we have attempted above. What follows, therefore, will be reduced to a thin skeleton of bibliographic orientations.

Upon which principles should the *classes of words* be defined and designated? Such was the second item on the agenda of the second session of the seventh International Congress of Linguists (London, 1952).⁷⁵ Agreement was practically unanimous on definition according to syntactic function.⁷⁶ O. Funke presented a remarkable comprehensive outline.⁷⁷

At the same congress the question (A4) of numerals and classifiers was examined.⁷⁸ To the latter could have been added that of the grammatical *gender*, concerning which there exists an enormous (and indigestible) compilation, practically summarizing all that has been written up to 1925;⁷⁹ but ideas on this subject have evolved considerably since 35 years ago, and an up-to-date comprehensive study would be necessary.⁸⁰ As to *verbal categories*, there does not seem to exist any synthesis analogous to that of Royen.⁸¹ The category of *person* has been studied by Paul Forchheimer.⁸²

The *systems of cases* have been the subject of classical works by Hjelmslev and Jakobson.⁸³ *Negation* was discussed fairly thoroughly at the seventh Congress of Linguists.⁸⁴ The expression of notions of *gradation*, *totality*, *ending-point*, *indication*, have been discussed in four published chapters which should have been parts of a series of studies on the 'logical and psychological bases of language', undertaken under the direction of Edward Sapir for the International Auxiliary Language Association, at a time when, according to Carroll,⁸⁵ he seemed to desire to use

'all the data available concerning scientific linguistics' for the purpose of creating a 'completely synthetic language', and, of which, unfortunately, the remainder never saw the light of day.⁸⁶

From the point of view of the present report a close study should also be made of *derivation* phenomena.⁸⁷

319 *Related problems*

We shall indicate very quickly here a few of the trends concerning general problems, which should be dealt with in a later detailed study, but which we can only touch upon in this report.

Problems of semantics. These cannot, as it has been observed, be neglected, even when they relate to grammar (morphology and syntax).⁸⁸ They become fundamental when they relate to the lexicon. It should be noted here that present-day lexicology is also tending to become structuralistic; no longer limiting itself to enumerating the words in an arbitrary alphabetical order, it seeks to classify them ideologically and to base its research on the study of conceptual systems.⁸⁹

Relationships between language and logic, language and thought, language and culture. We have grouped together these three subjects of discussion, since they can hardly be separated very practically and, furthermore, they have been, in recent years, the subject of passionate controversies—not very fruitful, however, in the absence of sufficiently exhaustive statements of facts, on a part of which each one claims to be based. A very prolific literature has been published, which we shall not attempt to analyse: we shall merely indicate a few guiding threads.

The question of the relationships between logic and language came up at the seventh International Congress of Linguists, in a very disputable form, however—which was perhaps the reason for the disappointing character of the discussion.⁹⁰ A few general outlines can be found on the subject, which do not appear to us to be very satisfactory;⁹¹ certain studies of detail are of greater interest.⁹²

The relationships between categories determined by the languages and the 'world views' (*Weltanschauungen*) of those who use them, have recently been the subject of a renewal of interest, especially in the United States, following the research undertaken by Benjamin Lee Whorf, on the basis of a hypothesis already formulated by Sapir (notably in the 1924 article mentioned in Note 42) on the influence of linguistic structures on conceptual systems.⁹³ The 'Sapir-Whorf hypothesis' has found a certain number of defenders⁹⁴ and many antagonists.⁹⁵ A certain number

of Whorf's formulae are assuredly bold in character; he has, however, the merit of posing clearly and strongly a problem of foremost importance, of which the solution, indeed, will have to await much more general and systematic investigations than those at present available.⁹⁶

Here, however, we immediately run into questions concerning the psychology of language,⁹⁷ its psychopathology⁹⁸ and to the 'primitive mentality'.⁹⁹ It is impossible for us to dwell on this subject here; we shall limit ourselves to indicating a few references—but these fields of study should also be explored from the point of view of our subject.

Types of languages. The attention devoted by historians of language of the nineteenth century (and the first decades of the twentieth) to the genealogical classification of languages has for a long time assigned a secondary place to their classification by types: in linguistics one is therefore faced by a situation inverse to that experienced in the natural sciences, of which the descriptive classifications had been established and perfected well before attention was paid to evolutionary classifications.¹⁰⁰ This situation is in the process of changing,¹⁰¹ but we are still far from possessing a satisfactory typology of the languages of the world, except perhaps from the phonological point of view.¹⁰² One may, moreover, be permitted to think that, in the more complex fields of grammar and vocabulary, only the intervention of the techniques of analytical codification and of treatment by mathematical methods, will lead to valid results; here there are research programmes of the greatest interest, made possible now by the evolution of automatic machines for the treatment of information.¹⁰³

Evolution of language considered as a more or less effective mechanism for the transmission of information. All of this brings us to what may be the most important question (from the standpoint of the creation of a 'unified scientific language'): if language is considered as a more or less efficient technique for 'communication' among men, what evolution can be discerned in this connexion from the point of view of its lesser or greater effectiveness? This question—although not raised exactly in the above terms—has long been the subject of debates between linguists concerning progress (or decadence or stagnation) in the evolution of languages.¹⁰⁴

It might be of interest here to link the evolution of the recent occidental trend of linguistic thought with that of contemporary philosophy, which, in a certain sense, has been dominated by what G. Friedmann refers to, in a remarkable book, as 'the crisis of

progress'. During the optimistic periods, which were still those of Meillet (and of Jespersen), the idea that evident progress could be discerned in the history of languages did not raise any doubt.¹⁰⁵

With Sapir, in 1921, things changed. He does observe three evolutionary tendencies for the English language: 'to eliminate the distinction between the subject and the object' (which is 'a recent chapter in the continuous simplification of the old Indo-European system of syntactic cases');¹⁰⁶ 'a tendency toward a fixed position in the phrase, a position determined by the syntactic relationship of words';¹⁰⁷ 'a tendency which carries language toward the unvarying word'.¹⁰⁸ But he draws no general conclusion on the fact that this would represent some sort of progress, nor on what could correspond to such tendencies outside the field of the English language. And, towards the end of his book, he declares that 'all attempts at linking certain morphological types to certain stages of culture . . . are in vain . . . types of languages, simple or complex . . . will be encountered at any cultural level'.¹⁰⁹ However, it can be observed here and there that Sapir has certain ideas concerning a linguistic type, to which he refers in his typology as 'simple languages with pure syntactic relationships', and which, he writes, 'offer an example of the most perfect linguistic expression'.¹¹⁰

If we now take the most recent American work, Hockett's manual, one notices a systematic rejection of anything that might bear resemblance to a judgement of value, which results in a sort of levelling doctrine on the one hand, and to the negation of any concept of direction of evolution, on the other.¹¹¹ In Whatmough one finds thought with finer shades of meaning, recalling at times that of Meillet.¹¹² In Whorf there are formulae which lead one to suppose that he considered certain so-called primitive languages to be more appropriate for a scientific description of the universe than what he termed the 'Standard Average European';¹¹³ in fact, however, what he wanted most was to combat the belief in certain 'eminent' qualities of such and such an Indo-European language, and in particular of English.¹¹⁴ Greenberg devotes an interesting chapter of his *Essays in Linguistics* to an examination of the evolutionary tendencies of language.¹¹⁵

We shall cover very briefly European research in this field: which does not seem too considerable. In J. Vendryes' lecture over the Radiodiffusion Française in 1951, there is somewhat of a return to the Romance theory (from Bopp to Max Müller) of the 'degeneration' of language.¹¹⁶ In 1922, Marcel Cohen adhered approximately to Meillet's theories;¹¹⁷ he has since been observed to adopt a much more reserved attitude.¹¹⁸

If an article by R. L'Hermitte¹¹⁹ is to be believed, the Soviet linguists are at present raising objections to Jespersen's idea of the

existence of a general law of progress in language. One should, however, bear in mind the possibility of differences of opinion and abstain from any conclusion in the absence of a more complete study, which we have been unable to undertake.¹²⁰

It may be observed at this point that there are objective reasons for the prudence observed today by linguists towards the theories of progress in language; and this is due mainly to newly discovered, or better studied, facts, which have tended to invalidate the older theories: e.g., the reconstruction of the laryngeals of the very ancient Indo-European languages,¹²¹ or the discovery in Dravidian languages of an evolution in a reverse direction to that of the Indo-European languages during the same period,¹²² and the taking into consideration of a 'linguistic horizon' much vaster than the Indo-European domain where Meillet or Jespersen were practically entrenched (with a few Hamito-Semitic or Finno-Ugrian incursions).

One question which relates closely to that of the evolution (and of progress) of language is that concerning the *factors of differentiation or unification* affecting languages.¹²³

32 Categories and relationships in 'artificial' languages

The subject which we have just broached leads rather naturally to an examination of the so-called 'artificial' languages designed to serve international relations. However, before considering them, it should be noted that their study is related fairly closely to that of the 'pidgin' and the creole languages. In fact, Hockett considers that the artificial languages (a qualificative should be added: *a posteriori*, in accordance with the usual terminology since Couturat; although we consider this term to be somewhat inappropriate, we shall preserve it, for lack of a better one) can be considered as varieties of pidgins,¹²⁴ and this thesis may be retained, as a whole.

The advantage of such a concept is in linking attempts at consciously 'constructed' international languages to a sort of typological family of much more numerous languages, also born of intercommunication requirements between people of very varying cultures and speaking languages of different structures.¹²⁵

A rather abundant bibliography is available concerning 'artificial' languages (consciously constructed for the purpose of intercommunication), but their scientific study has fundamentally made scarcely any progress; the available literature is mainly polemical and propagandistic.¹²⁶

Sapir was one of the few linguists interested in the construction of an international language in a truly objective spirit;¹²⁷

Jespersen was not objective, having been himself the creator of 'Novial';¹²⁸ A. Martinet collaborated in the editing of *Interlingua* from the 'naturalistic' point of view adopted by IALA¹²⁹ after the death of Sapir. Whatmough devotes a few pages to the question.¹³⁰ Logicians—or rather 'logisticians' as the promoters of mathematical logic were named in the 1900s—have tackled this field more or less successfully.¹³¹

Couturat divided the study of 'attempts toward the creation of a universal language' into two well-defined periods: that of the 'a priori' projects, i.e. those which do not derive their vocabulary from an existing stock, and that of the 'a posteriori' projects, i.e. basing their vocabulary on the 'common fund' of such or such a group of languages. This was, in his mind, to some extent a division between prehistory and (true) history, for he held nothing but scorn for the 'a priorists' (the term itself is somewhat belittling). However, since then, these have been given their due in the homage to Wilkins.¹³² The 'a priori' type is by no means extinct, and is represented among others, today, by Stuart C. Dodd's 'Tilp',¹³³ and Herbert E. Salzer's 'Geo'.¹³⁴

If one were to rewrite Couturat's book today (which, in addition, would meet a need), it is likely that one would observe in the IALA *Interlingua*, as in other such contemporary projects,¹³⁵ the continuation of an evolutionary 'lineage', which began with Schleyer's *Volapük* (1880), and of which the fundamental purpose is to attempt to constitute a 'universal language' on the basis of a sort of Germanico-Latin *lingua franca* (with a slight injection of Slav, for Zamenhof's *Esperanto*), retaining from among the grammatical characteristics more or less frequently represented in these groups of languages those which the author (or authors) of each new project considered as essential, or as more developed, or as more general. This would be achieved by elaborating a basic vocabulary, either by searching for 'common' roots (Schleyer), or by borrowing from various languages in the group, on a more or less equalitarian basis (Zamenhof), or by trying to 'go back through the stages of phonetic evolution' to find a sort of common *Urform* (IALA); and by applying a derivation method inspired by these languages and seeking only to 'rationalize' their affixes, according to various procedures.¹³⁶

Another trend seems to be distinguishable, of which the origin is found in Peano: his main characteristic is to seek simpler and more effective solutions with regard to grammatical structure.¹³⁷ Its second representative is Ogden's and Richard's *Basic English*; the objective of grammatical simplicity is sought here, however, by means of an entirely different method—a reduction in the number of verbs to 18 auxiliary verbs and some 300 verbal derivatives of substantives.¹³⁸

Lancelot Hogben's Interglossa is a sort of a cross between Peano's Interlingua and Basic English; it borrows from the latter its method of vocabulary reduction (a combination of a small number of 'verboids' with a large class of 'amplifiers', of which each can be name, adjective, or adverb). It is inspired a great deal by Peano for the remainder of its grammar and for the formation of its vocabulary, with this small difference, that where Peano used Latin words (rendered invariable), Hogben prefers to draw from the Greek stock.¹³⁹

A few words remain to be said concerning another trend, which doubtless appeared a long while ago (with Wilkins in 1668), but which later underwent a long eclipse: that of resorting to a symbolism not alphabetical, but ideographic. Today, if we are not mistaken, the first to have advocated an ideography as an auxiliary international language was the Pole, Edmund Erdman.¹⁴⁰ Holmstrom's report mentions (expressing strong doubts as to their chances of adoption) the similar ideas of the Englishman R. H. Richens and of the American Stanley Gerr.¹⁴¹ But the only one who, to our knowledge, had truly attempted to constitute on this basis a complete and practical system of expression is the Austrian chemist C. K. Bliss, now settled in Australia. His extensive work in three volumes¹⁴² is a considerable undertaking, which deserves much more attention than it has received so far.¹⁴³ Bliss clearly attempted to go further than Ogden in the direction of a reduction of the basic vocabulary, without, however, reaching Dodd's number 10: he uses a hundred or so elemental symbols. We lack the space, in this report, to outline in detail this extremely ingenious system, worthy of detailed study—whatever opinion may be formulated concerning the principle itself of an ideographical language conceived merely as a universal writing tool. Much less shall we attempt to discuss it here.¹⁴⁴

4 Problems of symbolization and notation

We had first intended here to examine the problems presented by the symbolization of general categories and relationships, on which we had previously published some preliminary studies, which are now, however, partly superseded.¹

We have observed, however, that this question needs now to be entirely re-evaluated on new bases: taking into account the advances made recently in the mathematical study of language,² in phonological research³ and perhaps even on the 'symbolism of sounds'.⁴

We therefore deemed it preferable not to enter into a discussion of this subject without having previously devoted to it a more comprehensive study, which we have not at present the leisure to undertake; we hope, therefore, that we will be forgiven for limiting this chapter to the simplest statement. It appears certain to us, however, that a collaboration of specialists in these matters, under their various aspects, will be necessary if, in the future, certain errors are to be avoided, which have occurred in the past in the symbolization and notation of classifications, as well as in that of 'universal languages'.⁵

Having reached the end of this study, we are under the impression that it has presented more problems than it has been possible to solve; the most concrete conclusion which can be drawn in this report is that a tremendous task remains to be undertaken: in research, assembling and coding of necessary data, the establishment of hypotheses to be subjected later to verification by calculation and experiment. We hope that, in the years to come, numerous research groups, gathering specialists from various fields of knowledge, will devote themselves to this task; that an effective system of exchange of information and of co-operation will be established between these groups, on an international scale.

Notes

The italic figures in the margin refer to page numbers

Introduction

- 10 1. This fundamental difference has been expressed in a number of more or less satisfactory formulae: 'multiple coordinate' classification systems, as opposed to 'linear' (Desaubiaux, Cordonnier) systems; 'analytico-synthetic' classification systems as distinguished from 'enumerative' systems (RANGANATHAN; see, in particular, *Colon classification*, 5th ed., vol. I, p. 1-12, Madras, 1957). We have adopted here the terminology used by Calvin N. MOOERS, 'A mathematical theory of language symbols in retrieval'. In: *International Conference on Scientific Information*, vol. 2, p. 1346. Washington, National Academy of Sciences, 1959.
- 10 2. Conference at the Union Française des Organismes de Documentation, held on 23 December 1943; 'Classification et classement' *Bulletin d'information scientifique et technique*, no. 6, May 1944, p. 21, 28, 51.
- 10 3. In the book edited by J. H. SHERA and Margaret E. EGAN, *Bibliographic organization*, p. 70-1, Chicago University Press, 1951.
- 10 4. L. I. GUTENMAKHER. An article in the *Vestnik Akad. Nauk*, 22 (1952), translated by the British Ministry of Supply, Technical

- Information Bureau, document TIB/T4271, January 1954, under the title 'On the question of mechanical handling of scientific information'.
- 10 5. 'The need for a faceted classification as the basis of all methods of information retrieval', document IACDocTerPAS 320/5515, 26 May 1955; reprinted in the *Proceedings of the International Study Conference on Classification for Information Retrieval held at . . . Dorking*, appendix 2 (cf. p. 146) London, Aslib, 1957.
- 10 6. Unesco document NS/115/4405, 9 April 1954.
- 11 7. 'Methods applicable to retrospective scientific document research', Unesco document 320/5601, p. 31. During its second meeting in London (17-19 November 1955), the IACDocTerPAS, following the presentation of a proposal by Mortimer Taube concerning the organization of an international conference on documentation research under the auspices of the American Documentation Institute, had declared itself in favour of the participation of Unesco in such a conference (which became the International Conference on Scientific Information, held in Washington in November 1958) and had recommended that 'M. de Grolier's report be forwarded to ADI as a preliminary document for the conference' (Unesco document NS/135, 30 April 1956, resolution N [3], p. 9).
- 11 8. *Proceedings* (see Note 5), p. 112. The conference consequently endorsed to some extent an idea which had been presented in our third report to the FID/CA Committee (General Theory of Classification), wherein we advocated that, in order to establish a new general classification system, the most important step would perhaps be to 'concentrate at the beginning on a systematization of the more basic "categories" applicable to the largest number of fields, e.g. logic, mathematics, general concepts of movement, form' (p. 9 of the typed document, 23 June 1955, reproduced in the *Revue de la documentation*, no. 1, vol. 23, 1956, p. 17).
- 12 9. Unesco document NS/148, 13 December 1957, resolution III/I, p. 4.
- 12 10. Yehoshua BAR-HILLEL, 'The present state of research on mechanical translation', *American documentation*, no. 2, October 1951, p. 234-5, (published in 1953).
- 12 11. See references no. 9 and no. 23 (June and September 1952) in the bibliography prepared by Victor H. YNGVE for *Mechanical translation*, vol. 1, nos. 1-3, and reproduced in the book published by William N. LOCKE and A. Donald BOOTH, *Machine translation of languages* p. 228 and 231, New York: Wiley, 1955.
- 12 12. *Ibid.*, reference no. 17, June 1952.
- 12 13. *Ibid.*, reference no. 30, May 1953.
- 13 14. 'Studies already being carried on in various countries, covering the logic of syntax, universal grammar, and semantic factoring, for instance, point toward the creation of a possible

inter-language or *machinese* into which concepts expressed in any given language could be translated, and from which these concepts could then be translated into another language.' BOOTH and LOCKE, 'Historical introduction', in the work mentioned in Note 11, p. 7. In the paragraph entitled 'Future prospects' of the same historical introduction, p. 12, the same concept is raised anew, in a more dubitative manner.

- 13 15. BOOTH, BRANDWOOD and CLEAVE, *Mechanical resolution of linguistic problems*, p. 292-5 London, Butterworths Scientific Publications, 1958. In his small popular treatise (the only one in French on the subject), Émile DELAVENAY gives a high priority to bilateral programmes (*La machine à traduire*, Paris, Presses Universitaires de France, *Que Sais-je?* series, 1959; cf. Chapter IV, p. 53-79; especially p. 57 and p. 78-9). It should be noted, however, that in a recently published article by A. D. BOOTH in French ('Progrès dans la traduction par machine', *Automatisme*, vol. IV, no. 3, March 1959, p. 97) he reverts to the idea of a 'meta-language of ideas', attributing to the very limited memory capacity of present machines the fact that they have not found practical application.
- 13 16. A. F. PARKER-RHODES, In: Allen KENT (ed.), *Information retrieval and machine translation*, vol. 2, reports nos. 51 and 39 New York, Interscience Publishers, 1960.
- 13 17. See the report prepared by Margaret MASTERMANN, R. M. NEEDHAM, and K. SPÄRCK JONES, 'The analogy between mechanical translation and library retrieval', *ICSI Proceedings*, vol. 2, p. 917-35.
- 13 18. Reports by O. S. AKHMANOVA (Institute of Linguistics, Academy of Sciences, Moscow) on 'Common machine languages as "auxiliary codes" for mediator languages' announced but not published; for N. D. ANDREEV on 'Universal code of science and machine language' see the publication quoted above, Note 16, report no. 49. Andreev is the chief of the experimental laboratory for mechanical translation at the University of Leningrad, concerning which Michel Corbé writes that its aim 'is to prepare the future' and that 'it is convinced that mechanical translation will be accomplished by the use of an artificial language which would serve at the same time as a language for automatic documentation' ('La traduction mécanique en USSR', *Automatisme*, vol. IV, no. 7-8, July-August 1959, p. 284). Corbé also mentions the Electronic Simulation Laboratory in Moscow as having devoted 'its greatest effort in this field to the construction of an artificial pivot-language to serve the purposes of both mechanical translation and automatic documentation' (p. 281). At the first Federal Conference on Mechanical Translation, held in Moscow, 15-21 May 1958, the problem of the 'intermediary language' was the object of a discussion which is summed up in the general report by Vladimir Andreevič USPENSKIJ on the results of the work of the section studying the algorithms of mechanical translation (cf.

p. 32-6 of the American translation of *Mašinnyj perevod i prikladnaja lingvistika*, no. 1, Moscow, 1959, document JPRS 925-D of the series 'Soviet developments in information processing and machine translation', since changed to 'Foreign developments . . .', Washington, US Joint Publications Research Service). Uspenskij mentions the research of Andreev and Mel'čuk, and criticizes the idea of Bel'skaja of using Russian as the intermediary language: 'There is no purpose in dragging along the inaccuracies of the Russian language; this is the reason for creating an intermediate language, to be rid of all the inaccuracies of the real languages' (p. 35); but he points out that there is some justification for Bel'skaja's point of view that the vocabulary of the intermediary (artificial) language must consist of concepts, and that 'it is not clear how far this can be done independently prior to the appearance of objective methods of investigating meanings'—methods which do not yet exist, but into which research must be done (p. 36).

- 13 19. We had broached this question briefly in our report to the First International Congress on Cybernetics (Namur, 1956), 'Cybernetics and documentation', *Actes du Congrès*, p. 195-6. Paris, Gauthier-Villars, 1958.
- 13 20. *Proceedings* (see Note 5), p. 113, point 9, d.
- 14 21. Such an undertaking could begin by a study of the duplicated document entitled 'Glossary and subject index' which is a revised and augmented edition of a glossary originally presented at the All-India Library Conference in 1953, but which is not sufficiently complete, nor sufficiently critical. The Indian secretariat of the FID/CA Committee at Insdoc, New Delhi, has published a duplicated document called 'Indian standard glossary for classification terms', which is based on the terminology of S.R. RANGANATHAN (*Annals of Library Science*, 5, 1958, p. 76-112).
- 14 22. See the summary of the discussion on the relationship between these two terms at Dorking, in the *Proceedings* (see Note 5) p. 89-90.
- 14 23. GUILBAUD. *La cybernétique*, p. 51, Paris, Presses Universitaires de France, 1954.
- 14 24. Joshua WHATMOUGH, *ICSI Proceedings*, vol. 2, p. 1031.
- 14 25. Jean PERROT. *La linguistique*, p. 9, Paris, Presses Universitaires de France, 1953.
- 14 26. Benoît MANDELBROT, Chapter III of his study entitled 'Linguistique statistique macrocosmique', in *Logique, langage et théorie de l'information* (by L. APOSTEL, B. MANDELBROT and A. MORF), fascicule III of the *Etudes d'épistémologie génétique* (published under the direction of Jean Piaget); cf. p. 54.
- 15 27. IBM report. 'Superimposed coding with the aid of randomizing squares for use in mechanical information searching systems', p. 3-5, Product Development Laboratory, Poughkeepsie, 1956.

- 15 28. Martin SCHEELE. 'Zur Frage der Beziehungen zwischen Klassifikation, Dokumentation und "Maschinensprache"', *Nachrichten für Dokumentation*, 9 (2), June 1958, p. 83. A similar restricted meaning is found in RANGANATHAN, 'Classification, coding and retrieval equipment', Unesco document NS/SL/3, 21 December 1951, p. 43: 'Coding may therefore be defined as follows: the translation of the names of specific subjects into written symbols suitable for presentation in the form of a series of perforations . . . etc., permitting mechanical retrieval.' See also the article by Geo. W. PATTERSON, 'What is a code?' *Communications of the Assn. for Computing Machinery (ACM)*, vol. 3, no. 6, June 1960.
- 15 29. 'A system of symbols for representing information and the rules for combining them', p. 6 of the duplicated document quoted in Note 21.
- 15 30. WILDHACK. *Proceedings, Dorking Conference*, p. 105. (See Note 5.)
- 15 31. D. J. FOSKETT, *ibid.*, p. 115. J. FARRADANE assimilates similarly facets and categories: 'groupings of similar concepts into categories of substantives, or facets' (*ibid.*, p. 65).
- 15 32. 'Glossary and subject index', p. 4. (See Note 21.)
- 15 33. 'Systematic subject indexing', *Journal of documentation*, vol. 9, no. 1, March 1953, p. 54; the same definition is picked up by B. C. VICKERY in 'Subject analysis for information retrieval', *ICSI Proceedings*, vol. 2, p. 858.
- 15 34. RANGANATHAN, *Colon classification* (see Note 1), p. I.60. The fundamental categories are: time, space, energy, matter, personality.
- 15 35. 'Classifying, indexing, coding', report to the Conference at Cleveland, quoted above (Note 16), vol. 1, p. 284.
- 16 36. A. LEROY and P. BRAFFORT. *Notice relative à l'élaboration d'un codage par phrases clés pour la programmation d'un système de sélection automatique des documents*, (note CEA, no. 278), p. 8, Saclay, Centre d'études nucléaires, 1959. The defined categories are objects or entities, properties, actions, conditions. See also P. BRAFFORT and A. LEROY 'Des mots-clés aux phrases-clés', *Bull. des Bibliothèques de France*, 4, Sept. 1959, p. 383-91.
- 16 37. This corresponds fairly closely to what we had indicated in our 'Concluding survey' of Dorking, concerning 'the necessity of establishing certain general categories . . . and of determining the general types of relationships between subjects which can be applied to special fields, in order to achieve a certain economy of methods and standardized procedures' (Point 6 of the 'Main points of agreement', p. 82 of the *Proceedings*). The picture of the 'sphere of semantemes' can be brought to mind, consisting of terms situated at various levels of abstraction and linked by a complex network of relationship lines, described by Charles L. BERNIER and Karl F. HEUMANN in 'Correlative Indexes, III:

Semantic relations among semantermes—the technical thesaurus', *American documentation*, vol. 8, July 1957, p. 213-4, and mentioned by Jesse H. SHERA in his inaugural address at Dorking 'Pattern, structure, and conceptualization in classification', *Proceedings*, p. 24. Earlier, in his 1953 article, mentioned in note 33 above, Vickery had indeed indicated that the three procedures using 'relationship prepositions', affixes or 'fundamental categories' are closely related and can be used fairly equally to express relationships between concepts (p. 51).

- 16 38. A first 'Codification des notions générales (logique-dialectique)' appeared in the form of an appendix in the manuscript of our report of 1952 for the Institut National des Techniques de la Documentation, subsequently published (1956) under the title *Théorie et pratique des classifications documentaires*, Paris, UFOD. This was a premature attempt, and we suppressed it in the published text.

General categories and relationships in general classification systems

- 17 1. 'L'étude des problèmes de classification documentaire sur le plan international', *Revue de la documentation*, 20, no. 3, September 1953, p. 112-13.
- 17 2. Especially at the Royal Library in Paris—which later became the Bibliothèque Nationale. See E. G. LEDOS, *Histoire des catalogues des livres imprimés de la Bibliothèque Nationale*, p. 37, 39, 40, Paris, Editions des Bibliothèques Nationales, 1938.
- 18 3. 'Classification on the shelves, with some account of the new scheme prepared for the Boston Athenaeum', *Library journal*, 4, 1879, p. 234-43.
- 18 4. Idem, *Expansive classification*, Boston, Cutter, 1891-93. Extended here to its 'sixth expansion'.
- 18 5. James Duff BROWN, *Subject classification*, London, 1906. The same author's 'Adjustable Classification' was published in his *Manual of library classification and shelf arrangement*, London, 1898. Earlier still, Brown had co-operated in the 'Classification of books for libraries in which readers are allowed access to the shelves', of which J. H. Quinn was the 'senior author', published in *The Library*, 7, 1895, p. 75-82.
- 18 6. B. C. VICKERY, *Classification and indexing in science*, p. 37, London, Butterworths Scientific Publications, 1958.
- 19 7. Paul OTLET, *Traité de documentation*, p. 382, Brussels, Editions Mundaneum, 1934.
- 20 8. D. KERVÉGANT, 'Développement de l'analyse des relations dans la CDU', memorandum duplicated in September 1957, and later published in: *Bulletin de l'Union Française des Organismes*

de Documentation, 26th year, no. 4, p. 1-5 (see p. 1); *Q. Bull. IAALD* 3, 1958, p. 111-6; and *Bull. des Bibliothèques de France*, 4, Nov. 1959, p. 502-3.

- 21 9. See VICKERY, *op. cit.*, (Note 6), p. 175. Properties are found notably in 54.03, 54.04, 615-011 and 678. Vickery also provides a number of examples of interactions, and more could be found very easily—such as the divisions of 625.04, which relate to ‘influence of the method of construction’ and to ‘atmospheric influences’ on highways, or of 625.036, concerning the ‘effects of the interplay of forces between highways and highway vehicles’.
- 23 10. These, until now, were characterized as ‘special’ divisions, belonging to certain specified classes: they still appear under this title on p. 24 of the trilingual edition of 1958. But the FID/CCC Committee, as supreme authority for the development of the UDC, decided recently to transfer gradually the hyphen divisions into ‘auxiliary tables’ (common); see memorandum F(Publ)59-75 of 24 July 1959. A very interesting commentary on the gradual extension of these ‘analytical divisions’ will be found in F. Donker Duyvis’ memorandum F58-63 (25 August 1958). We seize this opportunity to express our deep gratitude to the Honorary Secretary-General of the FID who was good enough, through an exchange of correspondence in April 1958, to guide us in the present study of general categories in connexion with the UDC.
- 26 11. See p. 359-62. But the 1937 German edition recommended (p. 348) the use of the alphabetical abbreviations of the Union Astronomique Internationale for the designation of constellations; the eighth international edition (in German) of 1958, Abt. 5, p. 44, does the same.
- 26 12. ‘L’étude des problèmes de classification documentaire sur le plan international’. *Revue de la documentation*, vol. 20, no. 3 (September 1953), p. 110-1. Attention is drawn to the fact that divisions (1/9) in their entirety are under general revision by Mr. Lloyd, at the British Standards Institution; the ‘correct’ use of analytic divisions (-0) and (-1), on the other hand, is examined by the General Secretariat of the FID.
- 26 13. Quoted by Jean BECQUEREL. *Exposé élémentaire de la théorie d’Einstein*, p. 21, Paris, Payot, 1922. The most important work on the idea of form is that of D’Arcy Wentworth THOMPSON on *Growth and form* (new ed.), London, Cambridge University Press, 1942. He defines form as ‘a ratio of magnitudes referred to direction in space’ (p. 78). Elsewhere he writes (p. 16) ‘the form of an object is a diagram of forces’.
- 28 14. The French translation is not accurate. The German title is ‘Äussere Form von Erzeugnissen. Halbzeug und zu bearbeitenden Fläche’ [External form of semi-finished products and the working surface]; and even the English title is restricted: ‘Form and shape of products and surfaces worked’.
- 28 15. For details, see the third international (German) edition of

1938, p. 815-6, with additions 1944, p. E6-7, and other complements in *Extensions and corrections to the UDC*, February 1954, under 621.9, derived from memo PE 433, which relates to continuous expansion headings. Unfortunately, we have been unable to refer to series 2 and 3 of the *Extensions and corrections*, covering the period 1953-57.

- 28 16. This .036 is applicable to the 624 divisions, notably, in so far as the form or shape of bridges is concerned 624.21.036; but forms of dams are direct divisions of 627.8 (627.824/828).
- 29 17. In recent PE notes it is indicated (PE 658, 17 Dec. 1958) that the -2/8 of 621 can be used—also, therefore the -4 under 684 furniture industry; but the same PE note provides new divisions 684.4.07 for the characteristics of elements of furniture, which are morphological divisions. Similarly, note PE 670 of 15 July 1959, provides extensions and modifications of 77.06 photographs pictures or prints according to their appearance, form, and size, instead of using -4. There would therefore be a considerable number of particularisms to be surmounted in order to obtain coherent common subdivisions by hyphen. This tends to prove that the UDC, beyond being a regular general classification, has become a collection of specialized classifications, developed side by side, but not truly integrated.
- 32 18. The existence will also be noted, under the analytical divisions at 616-039.1 (incidences of illnesses) of some temporal notions relating to seasonal illnesses, e.g., 616-039.18. The 616-039.3 relates to the length of illnesses. The 'before' and 'after' notions appear under the heading 616-089.1 pre- and post-operative care; in the 1951 German edition of class 6, 616-6 was headed complications. *consequences . . . prior condition* (the italics are ours), but only 'complications' appear in the 1958 trilingual abridged edition. In numerous places in class 6, of course, are found 'preliminary' (or exploratory) operations (by extension in 66.022) or 'completion' (by extension in 621.795). But we are surely here near the limits of purely temporal notions. In the analytical divisions of 66 industrial chemistry we find (note p. 682, 27 January 1960) the subdivisions 66-947 and 948 for operations according to their rapidity.
- 33 19. See our criticisms in the study (p. 111) mentioned earlier (Note 12).
- 36 20. The French text of the trilingual edition (but not the German nor the English texts) includes also evolution. Evolution is again found under natural theology, in 213.
- 36 21. Under 001.12 is found a heading 'character of things studied by science; beings, phenomena, facts, conditions'. Its division 001.123 production and cause of phenomena, coincides in part with 122; 001.124 actions, effects, and properties of phenomena, contains notions encountered also in 111.6 and 11.4; as to 001.126 application and utilization of phenomena, the latter corresponds to .004.14 in the common subdivisions of point of view.

- 36 22. That is -12 alternating movement; -13 continuous movement; -182 mobility: fixed, mobile, portable; -185 speed: slow, rapid; -193.2 machines in continuous operation; -193.4 in discontinuous operation. In 66 industrial chemistry is located a -932 division for continuous processes and another -934 for discontinuous processes.
- 37 23. In architecture, the notion of 'number of floors' is expressed by .011.26, curiously a subdivision of .01 aesthetics, theory. The notion of calculation is found in various places, e.g. in 628.22 calculation of the flow of sections of sewers; calculating machines are found under 681.14, in small precision mechanisms (681), being a part of a catch-all heading 68 miscellaneous industries and professions (continuation). Accounting is found at 657, but public accounting is under 336.126, commercial accounting from a legal standpoint in 347.719.
- 40 24. A few anomalies could be added. For instance tariffs were found at 621.317.8 under electrotechnics; they were cancelled in 1948 and referred back to 658.8.03; but under 69 construction, 'cost of construction, estimates, metering' was given under 69.003.12, thereby using divisions by .00. Why did electricians not use .003.13 sale price? or the builders 658.8.03? In an enumerative-type classification, such as the UDC, such inconsistencies are practically unavoidable, especially when the tables become too detailed, too copious, and too redundant.
- 40 25. See the German edition, 1958 Abt. 5, p. 243; there, in fact, biometry is found at .087.1, as a division of a .087 Messung und Aufzeichnung von Erscheinungen [Measurement and recording of phenomena]. The English text in the trilingual edition of 1958 shows under 578.087 'biometry, measurements', but the German and French texts show only 'Biometrie' and 'biométrie'.
- 43 26. See VICKERY's criticism, *Classification and indexing in science*, p. 131-4.
- 43 27. 'The problem of classification for bibliography, and a proposal', published in the *Revue de la documentation*, and later as publication FID 231, Classification studies, V, of the Comparative Classification Committee (which subsequently became FID/CA).
- 44 28. See J. Edwin HOLMSTROM, *Facts, files and action in business and public affairs*, vol. 2, p. 46-8, p. 199, and references, London, Chapman & Hall, 1953; also E. J. COATES *Subject catalogues; headings and structure*, p. 39-43, London, Library Association, 1960.
- 44 29. We shall not examine here the criticisms which may be directed against the CC on this point. See VICKERY, *Classification and indexing . . .*, p. 44, and our article 'Après Dorking', *Revue de la documentation*, vol. 25, no. 1, 1958, p. 12.
- 45 30. ' . . . a more or less all-comprehensive relation which is not merely any one of the other relations, listed in the Schedule', *Colon Classification*, 5th ed., p. 1.77.

- 45 31. His work *Classification and indexing in Science* (hereinafter abbreviated as *CIS*, p. 42.
- 45 32. The fifth edition, in fact, enumerates thirty-nine. But the z , 'generalia' is of purely bibliographical interest. 'By "generalia";' explains Ranganathan (p. 1.81), 'is meant a class into which should go any publication which deals with several subjects which cannot go into any other single main class.' He has, in fact included therein subjects such as $z41$, sinology, or zG 'Gandhiana', which really, should be located in class V; history, if the latter is taken in a rather broad sense. On the other hand, classes Γ physical sciences, μ humanities and social sciences, ν humanities, and Σ social sciences, are, actually, groupings of main classes, having no individual existence of their own. Had the CC known the equivalent of the / symbol of the UDC Γ could have been C/F, μ A/Z, ν N/S, and Σ T/Z. We must admit that we do not clearly see the differences between β and B; what the CC means by 'mathematical sciences', as distinguished from mathematics, is also not clear. As to 1 'universe of knowledge' and A 'Science', these are approximately the equivalent of A/Z.
- 46 33. The fifth edition states (p. 1.111) that class M '... is a miscellaneous class. Indeed it may be called a hold-all class. It should accommodate all the arts and crafts, in fact all the applications of sciences, which are not provided for in the other (MC) (= main classes)'. It warns us later that chapter M as it now stands 'is only brief and illustrative' (it is indeed contained on two pages) and adds: 'The number of subjects is infinite. Further, it is only in recent years that any appreciable literature has begun to grow in these subjects.' Here, the origin of the CC in an economically underdeveloped country probably makes itself felt. It will be noted elsewhere that 'book science' and journalism are also main classes under 3 and 4, which appear to duplicate M1 and M2.
- 51 34. In the schedules, p. 2.88, are found 'Foci in [2M]' for pharmacognosy (sic), but this must be in the nature of a typographical error, since the rules, p. 1.108, show [2P] and indicate that the comma should be used as a symbol, which is the [P] symbol.
- 52 35. An 'illustrative schedule' of this 'second level space facet' appears in Ranganathan's fourth report to the FID/CA Committee, 17 May 1954, document F54-40, p. 5-6. There are no 'places of the ancient world' in the CC, which is an improvement on the UDC.
- 53 36. In all cases where an 'isolate' (a division) 'is capable of chronological formation or subdivision' or when its 'individualization' may depend 'conveniently on the period of origin or birth or on the year of first investigation or on the year of discovery or on the year of initiation or commencement or on the year of occurrence or on the year that may be definitely associated with the respective isolates in any other manner or for any other reason'. Rules, p. 1.66.

- 57 37. Notably with the 'mnemonic device', which attributes to each figure, for subsequent divisions to be created, a quasi-mystical 'meaning'. Cf. Rules, p. 1.70, which state that the MD is 'a powerful device which secures a high degree of autonomy for individual classifiers. For this reason it can be handled only by classifiers who are highly attuned to the insight—I would even say, spiritual insight and the identity of patterns at seminal depths—or intuition which enables one to perceive directly—i.e., unassisted by the ordinary senses—the very depths of entities, the depths at which are disclosed the identity of patterns in spite of the differences in their manifestations at the level of sensual perception.' We must admit, as far as we are concerned, that we do not claim to possess such quasi-supernatural clairvoyance, permitting for example the assignment of the figure 5 to energy to foreigners, to emotion, to leaves, to esthetics, to woman, to sex, to crime, and to 'all other correlates to the above' (there is also the ocean, the 'public controlled plan', etc.). Such seminal depths, alas, remain inaccessible to us.
- 57 38. On the classification of the State Central Book Chamber, see 'Vsesojuznaja Kniznaja Palata', *Klassifikacija literatury v organakh gosudarstvennoj registracionno-učetnoj bibliografii*, 2nd ed., 1955; its schedule is given in P. L. HORECKY, *Libraries and bibliographic centers in the Soviet Union*, p. 193, Indiana University Publications, 1959. This latter book gives also class 3 of the 'classification for small libraries' (p. 194-6) and the main classes of the new classification of the Lenin Library (p. 197). The Library Commission of the USSR Academy of Sciences gave its approval on 11 June 1959 to a new library classification, on which see *ASLIB proc.* 12(2), p. 47-8 (translation from *Sovetskaja bibliografija* 3(55), p. 102-4, 1959) and 'The approval of basic order of the Soviet Library classification', *LLU Translations Bulletin* 2(2), February 1960, p. 129-33. The same *LLU Translations Bulletin* 1, January 1959, p. 5-7, gave a translation of an article by I. V. MOLODCEV, 'Problems of library classification', which appeared in *Vestnik Akad. Nauk SSSR* 1958(9), p. 122-3. A note in the *FID News Bulletin* 10(12), 15 December 1960, p. 33, tells us that: 'Work on the Soviet library classification has been considerably intensified; the chemistry division has already been released for printing, whereas the biology and agriculture divisions are under preparation.' Meanwhile the All-Union Book Chamber published at the end of 1960 its *Tablicy Universal'noj Desjatičnoj Klassifikacii*, which was 'not registered as an FID publication' (*FID News Bulletin* 10 (11), 15 November 1960, p. 31). Without doubt they incorporate at least some of the ideas which Z. N. Ambarcumjan presented to the Conference on Decimal Classification held in Budapest in July 1959, and which *Library Science Abstracts* summarize (abstract no. 9781) from an article in the *Knihovnik* of Prague (vol. 4, 1959, p. 264-5), which we have not seen. Neither do we know the Polish article by Maria Rojsowa, 'O projekcie radzieckiej klasyfikacji bibliotecznej', *Bibliotekarz*

- 22 (7), 1955, p. 193-8, except from the summary by Ida Forest in the *Bull. des Bibls. de France* 1 (1), January 1956, p. 60-62.
- 58 39. 'Méthodes nouvelles de documentation', Circular series X, no. 5 bis, 15 February 1946, of the ITBTP; cf. p. 14. We quote from this publication which is probably the most accessible one at this time, 'Classification et classement' of May 1944, long since out of print. But Cordonnier's ideas had already been set forth, in approximately the same form, in this first publication (which had, besides, already been circulated as a duplicated document in October 1943).
- 58 40. *Ibid.*, p. 10.
- 58 41. *Ibid.*, p. 13.
- 58 42. Invented by the American TAYLOR (US Patent 1 165 465 in 1915), rediscovered by H. E. SOPER (US Patent 1 351 692 in 1920), then by Liber in France, and a fourth time by Batten in England (cf. his report to Aslib in 1947).
- 58 43. 'Classification terminologique et diffusion "Sélecto"', duplicated document, p. 3, 1957 (not published in the *Dorking Proceedings*). In a document 'Optimisation de l'organisation documentaire' (duplicated July 1959, then printed in the *Revue de la Documentation*, vol. 27 (7) Feb. 1960, p. 12-31) CORDONNIER presented, *inter alia*, a method of expressing relationships by 'particules de relation spécifiques et vectorielles' (p. 22-3 of printed text).
- 59 44. Henri CLAVIER, *Grille et profil encyclopédiques*, Paris, Hermann, 1942; *Méthodes et procédés de travail intellectuel*, Paris, UFOD, 1956; *Esquisse d'un panorama universel des connaissances*, Paris, UFOD, 1957. See also his article 'Renseigner, informer, documenter automatiquement', *Usine nouvelle*, spring number 1957, p. 241-8.
- 59 45. Louis GLANGEAUD, 'Classification scalaire des sciences de la matière et de leurs méthodes', *Revue générale des sciences*, vol. 62 (5-6), 1955, 30 p.
- 59 46. Joseph NEEDHAM, *Integrative levels: a revaluation of the idea of progress*, Herbert Spencer Lecture, Oxford University, 1937; *Time: the refreshing river*, London, Allen and Unwin, 1943, pp. 233-72. Glangeaud himself indicates this relationship, and recalls the names of WODGER and of E. GUYE (*L'évolution physicochimique*, Paris, 1922). The ideas presented by B. KEDROV ('La classification des sciences', report to the International Congress on Philosophy, at Zürich, 1954; text reviewed in *Voprosy filosofii*, 1955, 2, and translated in *Recherches soviétiques, philosophie*, cahier 1, 1956, p. 83-111) may to a certain extent bear some resemblance to the above general trend: cf. the passage on p. 106-7 of the French translation concerning 'degrees of development', but Kedrov relies still too heavily on the old Comtian conception of classification of the 'sciences', and not of natural 'objects'; despite the title of his article, Glangeaud essentially envisages

the latter. Kedrov attempts (p. 107-10) to move from a scientific classification 'of a ramified type' to a 'linear' classification for library uses; although the scheme arrived at in this manner is certainly superior to that of the Lenin Library of 1948-51, it is—as the author himself states—'artificially simplified'.

- 59 47. D. J. FOSKETT, 'The construction of a faceted classification for a special subject'. *ICSI Proceedings*, vol. 2, p. 64-5 and p. 69.
- 59 48. Barbara KYLE, *An examination of some of the problems involved in drafting general classifications and some proposals for their solution*, duplicated document. October 1958; and published in the *Revue de la documentation*, vol. 26 (1), 859, p. 17-21.
- 60 49. D. J. FOSKETT, Report to the ICSI (mentioned in Note 47), p. 871-2 and p. 875-6; cf. also 'Comments on fundamental categories in document classification', duplicated paper submitted for a symposium to the 1959 Delhi Library Conference; and 'Comparative classification', *Annals of library science*, vol. 6 (4), December 1959, p. 105-12; and 'Classification and integrative levels', duplicated document, October 1960.

General categories and the expression of relationships in specialized classification and coding systems

- 62 1. ASSOCIATION FRANÇAISE POUR L'ÉTUDE DES EAUX, *Tableau de la classification*, 8th ed., 1949, 1957.
- 62 2. Serge TCHAKHOTINE, *Organisation rationnelle de la recherche scientifique*, p. 59-63, (see also figure on p. 36), Paris, Hermann, 1938.
- 64 3. See INSTITUT INTERNATIONAL DE LA SOUDURE, *Table de classement de la documentation*, 1st ed., 1949; 2nd ed. as a supplement to no. 1/3-1958 of the *Bulletin de documentation de la soudure et des techniques connexes*. Upon his return to Poland, DOBROWOLSKI published a work on the theory of classification, *Budowa klasyfikacji*, Warsaw, Państwowe Wydawnictwa Techniczne, 1956, which is being translated into French.
- 64 4. See R. DESAUBLIAUX, *Le danger des classifications linéaires appliquées aux problèmes sociaux et économiques*, Paris, Bernard frères, 1943. Desaubliaux subsequently presented various papers on the 'interferential' classification system, as he named it, to the Documentation Commission of the National Committee of the French organization (CNFO), notably on 21 December 1945; brief duplicated analyses only of these have been published.
- 65 5. See publication *Fruits*, vol. 12, no. 9 bis, October 1957.
- 65 6. Classification elaborated by P. DIDELIN; it is still used for the order of headings of the 'Documentation technique' published in the *Annales de l'ITBTP*, but a detailed classification of the abstracts is provided in the UDC.

- 65 7. *Répertoire de la classification méthodique syllabique à coordonnées multiples*, 2nd ed., 1954 (document LCAN, no. 317). A classification system of the Cordonnier type was also applied for a few years in a legal and social documentation service of Electricité de France, but was later abolished. The Centre de Documentation de la Direction des Études et Recherches d'EDF, on the other hand, uses a detailed decimal notation classification system, compiled by M. GOURIO (*Plan de classification*, 1952); however, we do not find it necessary to analyse it here in detail, since it offers no technical innovation over the UDC. It uses the / in the same sense as the : of the latter, naming it 'indexation by correlation' permitting an 'indication of the relationship between a subject (a) and an incidental aspect, a complement, a point of view, precise details, etc. . . . (b).' It includes also a few 'decimalization tables', which consist of divisions common to a number of headings, corresponding to the UDC's analytical subdivisions (some of these will be found, for example, under E-1, L-3).
- 68 8. Cordonnier, in effect, had modified his system of notation in 1951, and outlined to the Comité National de l'Organisation Française his new ideas on the subject (cf. *CNOF*, 25th year, no. 4, 1951, p. 3-8, and nos. 5-6, p. 19-35). To our knowledge, this 'Cordonnier II' notation, utilizing segments of four letters (consonant-vowel-vowel-consonant) has not been adopted anywhere.
- 68 9. 'Quelques problèmes de codification posés par l'usage des machines en vue de la recherche de l'information et de la traduction des documents', report to the second International Congress on Cybernetics. Namur, 1958, p. 4; 'Les progrès et l'avenir du "langage classificatoire"', sixth report to the FID/CA Committee, p. 4. We have available a few duplicated copies of these reports, which will be sent upon request.
- 69 10. The address of the Centre de Psychologie Appliquée is 15 rue Henri Heine, Paris-16^e.
- 69 11. Robert PAGÈS, *Problèmes de classification culturelle et documentaire*, Paris. UFOD. 1955; this is the revised text of R. Pagès' paper for a diploma of the UFOD Technical Course in Documentation, written in 1948. 'Introduction à l'analyse codée: une technique documentaire en psychologie sociale et sciences humaines', *Recherches sociologiques*, 1955, 2, p. 55-66; 1956, 3, p. 65-76; reproduced in the *Bulletin de l'Union Française des Organismes de Documentation*, 1955, no. 3-4 and no. 7-8 'Introduction: l'établissement, l'utilisation et le classement des fiches', *Fiches de psychologie sociale*, 1958, no. 1, p. I-X. 'L'analyse codée, technique documentaire en psychologie sociale et en sciences humaines: présentation et résumé de la grammaire', *Chiffres*, 1959, no. 2, p. 103-22. The code itself has only appeared to date as a temporary duplicated edition, for the internal use of the laboratory.
- 69 12. In the last part of our sixth report to the FID/CA Committee, September 1958, p. 3-4.

- 70 13. We refer the reader, in connexion with this question of levels and calibres, which is not essential to the present discussion and concerns certain particular rules of symbolization, to the comments made by Pagès in the article published in *Chiffres*, p. 119-20. -5 to -3 correspond to the symbols + and / of the UDC; 4, 8 and ∞ (approximately) to the : symbol of the latter.
- 72 14. Citations quoted from the article in *Chiffres*, p. 106.
- 76 15. P. BRAFFORT, *Elaboration d'une classification pour le fichier matières du Service de Documentation du Commissariat à l'Énergie Atomique* CEA report no. 238, December 1953. The 2nd edition revised by J. Jung, is entitled *Classification alpha-numérique pour le fichier matières du Service de Documentation du Commissariat à l'Énergie Atomique*, CEA report no. 568, 1956; an *Additif* was published in June 1958 (by Mme Roule), for the chapter on biology and radiobiology.
- 76 16. Page 4. It may seem paradoxical to base a classification system, an intellectual undertaking, on the spatial disposition of filing equipment. Furthermore, the word 'imposes' is certainly an exaggeration.
- 76 17. CORDONNIER, Circular, series X no. 5 *bis* of the Institut Technique du Bâtiment et des Travaux Publics, p. 14. The following appear:
- | | |
|--|----|
| Infra-atomic scale (nuclear physics) | Ca |
| Atomic scale (elementary chemistry) | Ce |
| Molecular scale (complex chemistry, structures) | Ci |
| Average scale (solid states, liquids, gases) | Co |
| Universal scale (astronomy, geodesy, physical geography) | Cu |
- 77 18. Notably 'the obligation not to exceed four symbols for the composition of a number' that the original scheme still imposed.
- 78 19. Which would not have occurred if, in the construction of the 'square of 25 elements', it had been discovered that ABCDE of the first line should be assigned instead to the first column, and continuing in the same manner.
- 79 20. Charles W. HARGRAVE (ed.), *Subject headings used in the catalogs of the United States Atomic Energy Commission*, 1st ed., 1951; 2nd ed., 2 vols., 1958; document TID-5001.
- 79 21. Note 36 of our Introduction.
- 79 22. P. BRAFFORT, 'Stratégies optimales pour la recherche automatique des informations', report to the Frankfurt Conference of the ADIA (Automatic Documentation in Action), June 1959.
- 81 23. The rejection of the validity of the traditional categories of 'parts of speech' or 'grammatical species' has been accepted at least since Ferdinand BRUNOT (*La pensée et la langue*, 1922). But as so often occurs, the makers of codes for information retrieval are working here without using the results of previous efforts of linguists. It should be noted in this connexion that Note 278 attributes to the 'language' it defines 'a very strongly "agglutinating" structure close to Chinese' (p. 19). It is strange

- to see Chinese described as an 'agglutinating' language and inaccurate to compare a code including substantives, adjectives and adverbs to a natural language which does not know these distinctions.
- 81 24. See Louis COUTURAT, *La logique de Leibniz*, Paris, Alcan, 1901, and, recently, R. M. YOST, Jr., *Leibniz and philosophical analysis*, Berkeley, Los Angeles, University of California Press, 1954.
- 82 25. Y. LECERF and A. LEROY, 'Description d'un algorithme d'analyse documentaire', contributed to the first Congress of the Association Française de Calcul at Grenoble, September 1960, *Grisa report* no. 6; Y. LECERF and P. IHM, 'Eléments pour une grammaire générale des langues projectives'. *Grisa report* no. 1, April 1960; Yves LECERF, 'Une représentation algébrique de la structure des phrases dans diverses langues naturelles'. *Comptes-Rendus, Acad. des Sciences* 252 (2), p. 232-4 (session of 19 December 1960). Lecerf, in this 'représentation' reconciles the models of Chomsky and of Tesnière.
- 82 26. 'Provisional instructions for putting scientific texts into the form of diagrams', *Grisa report* no. 5, August 1960. J. C. Gardin considers, however, that Ruvinschii's inventory of relationships is still largely inspired by differentiations suggested by the structure of a natural language—French, and especially French prepositions—instead of being based purely on the distinctions needed by a documentary language, probably much less numerous. (Association Marc Bloch, Groupe d'Études Documentaires pour les Sciences Humaines (GEDSH), *report* no. 2, December 1960, p. 2, note 1.)
- 82 27. J. IUNG and N. VANDEPUTTE. 'Les données documentaires; leur manipulation', internal document of CEA, DOC-CEN/S-AFD-22, August 1960.
- 82 28. Jean-Claude GARDIN, 'Problèmes de la documentation', *Diogène* no. 11, p. 107-24, July 1955. *Le fichier mécanographique de l'outillage*, Beyruth, Institut Français d'Archéologie, 1956; 'Four codes for the description of artifacts', *American anthropologist*, vol. 60, no. 2, April 1958, p. 335-7; 'On the coding of geometrical shapes and other representations, with reference to archeological documents', *ICSI*, vol. E, p. 889-901; 'Trois projets d'analyse structurale intéressants des phénomènes humains', multilithed notes of the Centre Mécanographique de Documentation Archéologique, 1959; J. C. GARDIN and P. BRAFFORT, 'Mise en évidence et exploitation mathématique des structures dans les phénomènes humains', duplicated document (CN/SHM 5) for the Colloque sur l'Utilisation des Techniques de Calcul Automatique dans les Sciences Humaines, organized by the Commission de la République Française pour l'Éducation, la Science et la Culture, 16-18 June 1959; *Centre d'Analyse Documentaire pour l'Archéologie*, pamphlet published by the Centre National de la Recherche Scientifique, 1959; Association

Marc Bloch, Groupe d'Études Documentaires pour les Sciences Humaines (GEDSH), *Report* no. 1, July 1960, 'Compte-rendu des travaux effectués pendant le premier semestre 1960' (duplicated separately, then in *Grisa report* no. 3); *Report* no. 2, 'Compte-rendu des travaux effectués pendant le 2^{me} trimestre 1960' (duplicated separately, then in *Grisa report* no. 9, February 1961); 'Les applications de la mécanographie dans la documentation archéologique', *Bull. des bibls. de France* 5 (1-3), January-March 1960, p. 5-16; 'Où en est la documentation automatique?', *Automatisme* 5 (12), December 1960, p. 456-8; 'Programme d'études sémiologiques et documentaires (1961-65)', Maison des Sciences de l'Homme, duplicated document.

- 83 29. 'Trois projets . . .' introduction, p. 1 (see Note 28). It would be rather interesting to trace the historical antecedents of Gardin's descriptive method; one might, for example, find analogies in heraldry and in the Bertillon system; cf. J. E. HOLMSTROM, *Facts, files and action in business and public affairs*, p. 84, part II, London, Chapman & Hall, 1953.
- 83 30. There is no need here to enter into the details of the construction of the code, which include, notably, a conventional designation of 'elementary symbols' and 'operations' which result in identifying each ornament by 'code-words', systematically using the syllabical symbolization; see the article in the *American anthropologist*, p. 342-4. We shall merely point out that Gardin encountered the same problem as Leroy and Braffort regarding the reduction in the number of basic elements of the vocabulary: as he writes, each 'elementary symbol' could, with the exception of the dot, be considered as the result of various arrangements of points closely related to each other. However, such an analysis would lead to a much too lengthy symbolism, and the operations needed to derive the 'elementary symbols' from the point would be of no particular use at the succeeding levels (p. 344). This indicates, by the same token, that the distinction between elementary symbols and operations is merely an 'expedient' used for purposes of economy, in order to arrive at a convenient symbolization, and an efficient language.
- 84 31. It is by no means certain that declension is, effectively, the most common method; 600 million Chinese are ignorant of it. But this choice was, in fact, decided by the selection equipment employed: no other procedure could have been applicable to visual superimposed cards. There are also 'independent' elements (supplementary outside the phrase) and 'appositions' (supplementary beings within the field, or within a supplementary phrase).
- 84 32. Gardin explains why he 'enlarged the concepts' p. 349 of the *American anthropologist* and his comments in this connexion have a very general bearing for any information retrieval system; see his statements concerning 'diversely named phenotypes'. In the same article, p. 3, he also discusses very interestingly why

- he did not analyse the 'sailing boat' into a *boat* with a *mast* to which is *attached* a sheet of *cloth* used for the *propulsion* of the ship.
- 85 33. See note No. 3, under this heading, in 'Trois projets . . .' mentioned in Note 28 and 29.
- 86 34. J. TRIER, *Der deutsche Wortschatz im Sinnbezirk des Verstandes* [The German vocabulary in the sensory sphere of the intellect] Heidelberg, 1931; G. MATORÉ, *La méthode en lexicologie: domaine français*, Paris, Didier, 1950.
- 86 35. J. J. LUND and M. TAUBE, 'A non-expansive classification system: an introduction to period classification', *Library quarterly*, July 1937, p. 373-94. It could be considered that this work on the Koran is not unlike that of Father Busa for the establishment of a concordance of St. Thomas—but this analysis was philological in character, an index of *words*, whereas the analysis of the Koran is here purely *conceptual*.
- 86 36. 'Étude structurale des mythes', undertaken under the scientific authority of M. Levi-Strauss, see note No. 2 of 'Trois projets . . .'.
- 86 37. We refer the reader again to note No. 1 of 'Trois projets . . .': 'Mise en évidence et exploitation mathématiques des structures dans les phénomènes humains; étude topologique des formes de vases' for the exploitation of the inventory of the shapes of vases, described above, with a view to bringing out the structure of a collection of these vases, the 'centre of gravity' of this aggregate, and, finally, to permit its dynamic study (diffusion of the cultural type or evolution of a same family of vases, represented by the displacement of the 'centre of gravity'). This is no longer a coding problem, but a scientific analysis of phenomena which—we should mention—the analytical coding thus permits. Cf. also the Gardin and Braffort reports under the same title ('Mise en évidence . . .'), made to the Colloquium on the Use of Automatic Computation Techniques in the Human Sciences, Paris, 16-18 June 1959 (during the International Conference on Information Processing).
- 87 38. GEDSH, *Report* no. 1 (1st ed.) p. 1; *Grisa report* no. 3, p. 8.
- 87 39. *Idem*, *Report* no. 2 (1st ed.) p. 1; *Grisa report* no. 9, p. 3.
- 87 40. *Idem*, *Report* no. 1, p. 4-6, and table in Annexe II (1st ed.); *Grisa report* no. 3, p. 10-11 and Annexe II, p. 17. We have made some objections to Gardin's table in our study 'Points de vue rétrospectif et prospectif dans la théorie de la classification', to be published in the volume of essays edited by D. J. Foskett and B. I. Palmer in memory of Berwick Sayers.
- 87 41. *Idem*, *Report* no. 1 (1st ed.) p. 6; *Grisa report* no. 3, p. 12. For the reasons for cancelling the category 'process' as such, see *Report* no. 2 (1st ed.) p. 11; *Grisa report* no. 9, p. 10-11.
- 87 42. *Idem*, *Report* no. 1 (1st ed.), p. 10; of *Grisa report* no. 3, p. 15.

- 87 43. Idem, *Report no. 2* (1st ed.) p. 12; *Grisa report no. 9*, p. 11. The inventory of 'modes' is reserved for a later report. *Report no. 2* gives only some illustrations: 'descriptive', 'critical', 'systematic' mode, etc. (1st ed. p. 10, note 3; *Grisa report no. 9*, p. 10, note 2.)
- 87 44. Idem, *Report no. 2* (1st ed.) p. 13-16; *Grisa report no. 9*, p. 11-14.
- 88 45. Idem, *Report no. 2* (1st ed.) p. 17; *Grisa report no. 9*, p. 15.
- 88 46. Idem, *Report no. 2* (1st ed.) p. 20-34; *Grisa report no. 9*, p. 17-26.
- 88 47. Idem, *Report no. 2* (1st ed.) p. 34; *Grisa report no. 9*, p. 26.
- 89 48. Gardin insists on the practical character of the choice he has made, and suggests that experience will show whether the system of four relationships is 'still too rich, in view of the many syntactical implications which arise in practice from the terms of any discipline, or whether on the contrary it is too poor to "define" the groups of homogeneous propositions which documentary research aims at' (p. 34 of the 1st ed.; *Grisa report no. 9*, p. 26).
- 89 49. Idem, *Report 1st ed.*, p. 36; *Grisa report no. 9*, p. 28.
- 89 50. Idem, *Report 1st ed.*, p. 38-9; *Grisa report no. 9*, p. 29-30.
- 89 51. FRANÇOIS ISAMBERT, *Archives de Sociologie des Religions*, 1, January-June 1956, p. 141-58. The title of the bibliography in question is 'Sociology of religions; a trend report and bibliography', *Current Sociology*, 1956, 1, 87 p.
- 90 52. C. GROENEVELD, 'Problems of classification', *Revue de la documentation* 14, 1947, 99-115. The ideas expressed by Groeneveld have been adopted in part for some recent revisions in certain classes of the Universal Decimal Classification.
- 90 53. See G. J. KOELEWIJN, 'The possibilities of far-reaching mechanization of novelty search of the patent literature', *ICSI*, vol 2, p. 1071-96.
- 90 54. Same report by Koelewijn, and his paper to the ADIA Conference at Frankfurt, June 1959. 'Mechanische Auswahl nach Begriffen und deren Beziehungen zueinander' [Mechanical selection according to concepts and their mutual relationships].
- 91 55. *ICSI Proceedings*, vol. 21, p. 1080 and p. 1092-3, and J. DEKKER's report, 'Interfixing of descriptors for row-by-row coded punched card machine literature searching systems', International Conference for Standards on a Common Language..., Western Reserve University, September 1959, In: Allen KENT (ed.), *Information retrieval and machine translation*, vol. 1, p. 377-87, New York, Interscience, 1960. Dekker mentions therein an article by H. BOUMAN ('Enkele aspecten van de documentatie bij een bedrijfssoctrooiafdeling' [Some aspects of industrial patent classification: A Nider publication], *NIDER-publicatie*, 2nd series, no. 24, January 1959, p. 28-32) which describes the use of

- method of interfixes at the BPM (Dutch Shell); we have been unable to refer to this publication.
- 91 56. See the report on the 1951-58 research performed by the Working committee on mechanization of documentation [Arbeitsausschuss zur Mechanisierung der Dokumentation], by E. PIETSCH, *Nachrichten für Dokumentation*, 9. Jg. H.4, Dec. 1958, p. 190-3.
- 91 57. E. PIETSCH, 'Evaluation of mechanized documentation at the Gmelin Institute', In: R. S. CASEY, *et al.*, *Punched cards*, 2nd ed., p. 571-618, New York, 1958.
- 91 58. *Gmelin Handbuch der anorganischen Chemie, Systematik der Sachverhalte* [Gmelin Handbook of inorganic chemistry, systematic presentation of contents], Weinheim, Verlag Chemie, 1957.
- 91 59. I. KUBACH and W. MÜLLER, 'Klassifikation der Mineralien für die automatische Dokumentation' [Classification of Minerals for Automatic Documentation], *Fortschritte der Mineralogie*, 35, 1957, 38, 'Die Sphinxo-Sichtlochkartei für ein Teilgebiet der anorganischen Chemie' [The Sphinx peek-a-boo card file for parts of inorganic chemistry], *Nachrichten für Dokumentation* 8. Jg. H.3, p. 148-50. The classification system for the application of radiation to the food industries was elaborated by H. Lück. Deutsche Forschungsanstalt für Lebensmittelchemie, Munich [German Research Institute for Food Chemistry] and it is used for series C, section 02/1, 1959, Schriftenreihe des Referates Atomkernenergie-Dokumentation [Literature series on atomic energy documentation] of the Gmelin Institut. At the ADIA Conference, K. Schneider, chief of the Automatic Documentation Group of the Gmelin Institut, gave three demonstrations—two on a Univac Fac-Tronic I and one on an IBM 650—of the methods used.
- 91 60. Martin SCHEELE, *Die Lochkartenverfahren in Forschung und Dokumentation* [Punched card procedure in research and documentation], p. 80-81, Stuttgart, Schweizerbart, 1954.
- 92 61. Idem, 'Kodifizierung auf dem Gebiet der Biologie' [Codification in the field of biology], IBM, Germany, 9 September 1955, cf. p. 14. The 'questions' system is much less developed: mentioned only are extension fields of notions, one category of these—space, time, form, process—being of general scope. In a manual dedicated to Dr. Ernst Telschow, Martin Scheele has contributed a most interesting chapter on 'Die Bedeutung des Menschen als Voraussetzung wissenschaftlicher Systemkunde' [The significance of man as a prerequisite for scientific systematology] (In: B. RAJEWSKY and G. SCHREIBER (eds.), *Aus der deutschen Forschung der letzten Dezennien* [From German research during the last decades], p. 44-5, Stuttgart, Thieme, 1956). Paragraph 4 of this study is entitled 'Das Problem der Beobachtungsstufen' [The problem of observation levels]; through Lecomte Du Noüy it can be seen that the ideas of

- Charles Eugène Guye were those which influenced Scheele on this point. Like so many 'in the air' notions, this particular notion of levels of observation has a multiple paternity: the Swiss Guye, the Englishman Needham, the Frenchman Glan-gaud.
- 92 62. 'Neue Wege zur Einheit der Wissenschaften', *Studium Generale*, 8. Jg., 1955, H.7 [New ways to the unification of the sciences], (reprint); we incorporated a summary of this article in our fifth report to the FID/CA Committee, *Revue de la documentation*, vol. 25, no. 1, 1958, p. 14.
- 92 63. Idem, *50 Jahre Archiv für Hydrobiologie*, Stuttgart, Schweizerbart, 1958. The works published from 1906 to 1955 in this journal are indexed in it under a series of tables according to the standpoint of their methods, of 'limnological units' (sources, lakes, etc.), of physical and chemical factors, of 'geographical units', of groups of organisms; there are also summary (statistical) tables grouped in nine large (physiological) systems and their divisions.
- 92 64. O. NACKE, 'Concerning two principles for construction of a common machine language for medicine', Preprint, International Conference for Standards on a Common Language . . ., Western Reserve University, September 1959, In: Allen KENT (ed.), *Information retrieval and machine translation*, vol. 1, p. 503-13, New York, Interscience Publishers, 1960.
- 92 65. J. E. HOLMSTROM, *Facts, files, and action*, vol. 2, p. 70-83, London, 1953; see also the various references listed therein.
- 93 66. R. HAMMOND, 'The Brisch system of decimal classification in production', *Engineer*, 178, 1944, p. 109. The Brisch classification system could doubtless be tied in with F. W. Taylor's system. The Brisch method is used in continental Europe by Compagnie Parisienne d'Ingénieurs-Conseils (CPIC); an engineer of this company, S. Briès, has prepared several reports concerning the method, one of the clearest of which was made in his lecture on 4 May 1953 to the Syndicat Général des Industries Mécaniques et Transformatrices des Métaux ('La réduction des prix de revient par la classification 'Brisch-Copic', *Revue générale de mécanique*, July 1953; cf. notably p. 5-7 of the offprint). A Brisch classification system applying no longer to objects but to concepts has been evolved for the Service de Documentation de l'Agence Européenne de Productivité (*Système de classification*, project 320); it is much less interesting.
- 94 67. J. E. L. FARRADANE, 'A scientific theory of classification and indexing', *Journal of Documentation*, 6, 1950, p. 83-99; 8, 1952, p. 73-92 and 11, 1955, p. 187-201. A summary of Farradane's theory is given by E. J. COATES in his book *Subject catalogues: headings and structure*, p. 45-9, London, Library Association, 1960, and we have condensed the above information from this summary. See also a recent article by FARRADANE in *The indexer* 2, 1961, p. 127-33.

- 94 68. 'Some problems of management classification', *Annals of Library Science*, vol. 3, no. 4, December 1956, p. 105-8. We have included Miss Kyle in this group, although her name does not appear among those of its members as of 1 January 1955, but, nevertheless, she worked very closely with the CRG, and also participated very actively in the Dorking conference.
- 94 69. See the 'Bibliography of papers on classification and allied subjects, by members of the Classification Research Group', *Journal of Documentation*, 12(4), December 1956, p. 227-30. One of the most recent classifications worked out according to the methods of the CRG is that of A. S. Reid for the Nobel Division of Imperial Chemical Industries Ltd.: 'A faceted classification scheme for explosives technology', July 1960, duplicated.
- 95 70. D. J. FOSKETT, 'The Colon Classification', *Lib. Assn. Record*, December 1950, p. 450-5.
- 95 71. *Documents classification system: container manufacture and packaging*, Research Dept., Metal Box Co., Ltd., August 1957, duplicated. Summary in VICKERY, *CIS*, p. 158-61.
- 95 72. Summary in VICKERY, *CIS*, p. 156-8.
- 95 73. This classification has been subsequently adapted for the Library of the National College of Food Technology (Weybridge, Surrey, 1959; duplicated); here it has been integrated into a general scheme based on the classification of H. E. Bliss, and consequently with a modified notation. The facet for 'parts' has been re-named 'biology'.
- 95 74. See *Proceedings* of the Dorking conference, p. 115-36.
- 95 75. International Occupational Safety and Health Information Centre, *Guide to the card service and classification scheme*, Geneva, 1960. See also the article by FOSKETT, 'Documentation in occupational safety and health' *Revue de la Documentation* 27 (3), 1960, p. 102-7.
- 95 76. *Guide to the card service . . .*, op. cit., p. 7; article by FOSKETT, p. 103.
- 96 77. *Guide to the card service . . .*, op. cit., p. 7, explains that 'it seems clear that this division (into 5 categories) reveals the logical structure of the field considered; each of its elements in fact presupposes the existence of those which precede it, and influence the nature of those which follow it'—but this really seems a little doubtful, at least as far as the fifth is concerned, for if there were neither people nor industries there would also be no more risks.
- 97 78. B. C. VICARY, *Faceted classification: a guide to construction and use of special schemes*, London, Aslib, 1960.
- 97 79. *Ibid.*, p. 23-6 and p. 27-31.
- 97 80. *Ibid.*, p. 37: here are enumerated the relationships of influencing or effect, of comparison, of association, and of 'bias', a

- term taken from Ranganathan to indicate 'a subject studied from the point of view of another subject', e.g., statistics for economics, or physiological anatomy.
- 98 81. See also 'Towards a classification for social literature', *American Documentation*, vol. 9, no. 3, March 1958, p. 168-83.
- 99 82. 'Classification Research Group. Points for discussion on the general question of order', Duplicated.
- 100 83. 'Classification with peek-a-boo for indexing documents on aerodynamics; an experiment in retrieval', *Proceedings*, vol. 1, p. 771-801; the report contains an interesting evaluation of the system's efficiency, with an examination of the causes of failures noted. The latter are mostly due to indexing omissions (34 cases out of 55), deficiencies attributable to the classification appearing in 6 cases only.
- 100 84. P. R. P. CLARIDGE, 'Information handling in a large information system', *Proceedings, ICSI*, vol. 2, p. 1203-20; cf. fig. 1, p. 1205, which indicates the categories. Claridge's scheme has not been continued, having been finally judged to be too costly.
- 101 85. W. S. BATTEN, 'A punched card system of indexing to meet special requirements', *Aslib, Reports 22nd Conference*, p. 37-9, 1947.
- 101 86. G. M. DYSON, *A new notation and enumeration system for organic compounds*, London, Longmans, 1947, 2nd ed., 1949; 'Codification of chemical structures', *Research*, 2, 1949, p. 576; *Nature*, 165, 1950, p. 301; 'Studies in chemical documentation, III: mechanized documentation', *Chemistry and Industry*, 1954, no. 16, p. 440-9; 'Advances in classification', *Journal of Documentation*, vol. 11, March 1955, p. 12-18.
- 101 87. HANS SELYE and MIKLOS NADASKI, *Symbolic shorthand system (SSS) for physiology and medicine*, Montreal University, Institute of Experimental Medicine and Surgery.
- 104 88. 'Perhaps better systems can be devised on an entirely different basis but, in any event, it seems to us that the time has come when we have to develop a special script for medical topics, a set of Shorthand Symbols which correspond to the structure formulae and equations of chemistry, the symbols of algebra and the notes of music' (p. 61). We shall merely call attention here to Selye's remarkable observations on the problem of medical documentation in general and its organization: see p. 1-3 and p. 56-61 of his 'Introduction'.
- 104 89. ROBERT S. CASEY and JAMES W. PERRY, *Punched cards: their application to science and industry*, New York, Reinhold, 1951, 2nd ed., 1958; extensive bibliography. ISABEL L. TOWNER (comp.), *Classification schemes and subject headings list: loan collection of Special Libraries Association*, New York, SLA, 1951; the latter collection has since been transferred to Western Reserve University. ENGINEERING SOCIETIES LIBRARY, *Bibliography on filing*,

- classification, and indexing systems for engineering offices and libraries*, New York, Engineering Societies Library, 1954 (ESL Bibliography no. 9). Karl F. HEUMANN, 'Information theory in library and documentation activities', *Library of Congress Information Bulletin*, Appendix II, vol. 13, no. 39, 27 September 1954. Conference on the Practical Utilization of Recorded Knowledge, Western Reserve University, 28 December 1955; pre-conference paper no. 6, *Report of the committee on the inventory of methods and devices for analysis, storage and retrieval*, Cleveland, Western Reserve University (see the 57 references p. a-e); LOFTUS and ALLEN KENT, 'Appendix: an annotated bibliography', In: James W. PERRY, ALLEN KENT and Madeline M. BERRY, *Machine literature searching*, p. 135-47, New York, Interscience Publishers, 1956. NATIONAL SCIENCE FOUNDATION, Office of Scientific Information, *Non-conventional technical information systems in current use*, Washington, 1958, Idem, *Current research and development in scientific information* (4 numbers published: July 1957, April and October 1958, April 1959, preceded in August 1956 by a 'Preliminary report on research in progress in scientific documentation'). Charles P. BOURNE, *Bibliography on the mechanization of information retrieval*, Menlo Park, Stanford Research Institute, 1 February 1958, Supplement 1, 1 February 1959. Peter JAMES, *Bibliography and index: literature on information retrieval and machine translation*, New York Service Bureau Corp., September 1958; 2nd ed., June 1959. ALLEN KENT, 'Machine literature searching and translation; an analytical review' In: *Information retrieval and machine translation*, vol. 1, p. 13-236, New York, Interscience, 1960.
- 105 90. J. V. ATANASOFF and A. E. BRANDT, 'Application of punched-cards equipment to the analysis of complex spectra', *Journal of the Optical Society of America*, vol. 26, February 1936, p. 83-8. J. BERKSON, 'System of codification of medical diagnoses for application to punch cards', *American Journal of Public Health*, vol. 26, 1936, p. 606-12. F. SCHREIBER and A. NIELSON, 'Punch card code for classification of cranio-cerebral injuries', *Journal of the Michigan Medical Society*, vol. 37, 1938, p. 909-12. We have not examined these studies, cited here according to P. James.
- 105 91. 'Punched cards for a chemical bibliography', *Chemical Engineering News*, vol. 23, 1945, p. 1623-6.
- 105 92. D. E. H. FREAR, *Chemical Engineering News*, vol. 23, 1945, p. 2077.
- 105 93. The chemical code has been published: *A method of coding chemicals for correlation and classification*, Washington, CBCC, 1950; the biological code was duplicated only, in various temporary issues, the last in 1952: 'The detailed biological code of the Chemical-Biological Coordination Center', Washington, CBCC; an article by G. Congdon Wood gives a fairly detailed

description of the code: 'A detailed biology code for storing, retrieving and correlating chemical-biological data', *American Documentation*, vol. 8, no. 3, July 1957, p. 168-80.

- 106 94. See Leslie L. CLARK's report, 'Some computer techniques in the behavioral sciences', International Conference for Standards on a Common Language . . . , Western Reserve University, September 1959, published in *Information retrieval and machine translation*, op. cit., p. 445-66. On page 463 will be found a fragment of the OCM, and on page 465 a fragment of the OWC. The report indicates in an interesting manner the limitations of the method used to date, which does not permit certain types of research (see, in particular, p. 451-2); the ideas he outlines on the 'quantification' of data should be compared with those of Gardin, mentioned in Note 37.
- 106 95. Report presented by Calvin N. MOOERS to the American Chemical Society in September 1947; the basic outline is presently known as 'Zatocoding for punched cards', *Zator technical bulletin*, no. 30, 1950. A bibliography of Mooers' writings was published in May 1957, *Zator technical bulletin*, no. 105.
- 106 96. The 'Zatopleg' method for the coding of chemical compounds, devised by Mooers in 1947, served in 1956 the National Bureau of Standards in the experimentation in the automatic selection of steroids with an SEAC computer. See also 'Finding chemical information', *Zator technical bulletin* no. 64, 1951, and the important report 'Information retrieval on structured content', In: Colin CHERRY (ed.), *Information theory*, p. 121-34, London, Butterworths Scientific Publications, 1956 (papers read at the Third Symposium on the Theory of Information, London, 1955); see p. 124-8 for the 'interlocking sets'.
- 107 97. 'Application of random codes to the gathering of statistical information', Master's thesis, Massachusetts Institute of Technology, 1948, and *Zator technical bulletin* no. 31, 1949; 'Choice and coding in information retrieval systems', *Transactions of the Professional Group on Information Theory*, vol. PGIT-4, September 1954, p. 112-8; the report to the Third Symposium on the Theory of Information, London, 1955, mentioned in Note 96 above, the report to the ICSI in November 1958, mentioned in Note 1 of our 'Introduction'; 'Some mathematical fundamentals of the use of symbols in information retrieval', report to the Unesco International Conference on Information Processing, Paris, June 1959, Unesco document NS/ICIP/J.5.5.
- 107 98. In regard to the foregoing, the best outline was that presented by MOOERS at Aslib on 7 September 1955, 'Zatocoding and developments in information retrieval', *Aslib Proceedings*, vol. 8, no. 1, February 1956, p. 3-22. See, especially, a partial reproduction of a 'schedule of descriptors', (p. 14). See also Claude W. BRENNER's 'Experience in setting up and using the Zatocoding system', (Allied Research Associates, Inc.), reprinted as *Zator technical bulletin* no. 107, 1957, which furnishes examples of

- definitions of descriptors, and completes the 'schedule of descriptors' of the Aslib conference. During the past few months, Mooers has presented two very interesting outlines as to the future of 'information retrieval' methods—a term which he 'invented' in 1950 and which is mentioned for the first time in 1950 in his report, 'Information retrieval viewed as temporal signalling', to the International Congress of Mathematicians. 'The next twenty years in information retrieval: some goals and predictions', *Zator technical bulletin*, no. 121, March 1959, and his report, 'A system designed for a large-scale information retrieval center', to the ADIA Conference at Frankfurt-on-Main, June 1959. R. J. Solomonoff is also doing research at the Zator Company, and his studies on machines capable of learning may have important future repercussions on information retrieval methods: see his recent report to the WRU Conference, at Cleveland, September 1959: 'A progress report on machines to learn to translate languages and retrieve information' (*Information retrieval and machine translation*, op. cit., vol. 2, chap. 41), which provides previous bibliographical data.
- 107 99. A. G. GUY and A. H. GEISLER, 'A punch card filing system for metallurgical literature', *Metal progress*, vol. 52, no. 6, 1947, p. 993-1000.
- 107 100. *ASM-SLA metallurgical literature classification*, Cleveland, ASM, 1950; a revised edition was published in 1958. The Secretary of the Committee was Marjorie R. Hyslop. In Italy, the Associazione Italiana di Metallurgia had prepared, immediately following the war, a classification of metallurgy, which ran into eight successive editions; with a view to stimulating international co-operation, Professor Scortecchi, President of the Bibliography Committee of the AIM, was instrumental in having the ASM-SLA classification system adopted in Italy; the Italian translation, under the title *Classificazione bibliografica internazionale della metallurgia*, was published in 1955.
- 107 101. *Ibid.*, 'Part II—The punched card system', p. 5-15; the card was an adaptation of the card proposed by Guy and Geissler.
- 108 102. 'If as many as 10,000 items are to be indexed, the hand-punched card system becomes unwieldy', writes Marjorie R. HYSLOP in 'Forecast of an information center', *Metal Progress*, July 1958, p. 2 of the special printing. Our own experience would place the limit much lower (under 5,000 documents).
- 108 103. J. W. PERRY, 'Indexing and classifying results of chemical research in relation to punched card investigation', *Journal of Chemical Education*, vol. 24, 1947, p. 71-74; 'The use of punched cards in American libraries', *Aslib, Report on proceedings of the 22nd conference*, 1947, p. 40-50. It has been noted that J. W. Perry was the co-author, with R. S. Casey, of a report on the scientific and industrial applications of punched cards, published in 1951. On the Punched Card Committee of the ACS, see Perry's report in *Chemical and Engineering News*, vol. 27, 1949, p. 78.

- 108 104. James W. PERRY, Allen KENT and Madeline M. BERRY, *Machine literature searching*, New York, Interscience Publishers, 1956; we have provided a report on the foregoing in the Unesco (Department of Natural Sciences) document 320/7115, 15 January 1957.
- 108 105. Chapter V: 'Definition and systematization of terminology for code development', *American documentation*, vol. 5, no. 3, 1954; p. 19-26 of the book.
- 108 106. 'Chapter IV: Collection of terminology', *American documentation*, vol. 5, no. 2; p. 13-8 of the book.
- 109 107. *Machine literature searching*, op. cit., p. 22-3. The five categories in 'general fields' are subsequently replaced, at a third level of analysis, by a classification into nineteen classes of terms 'according to their use in limited fields of specialization or in various industries' (see p. 24).
- 109 108. The principle, however, was contained in 'Chapter VI: Class definition and code construction', *American Documentation*, vol. 5, no. 4, 1954; p. 27-33 of the book.
- 109 109. See Louis COUTURAT, *La logique de Leibniz*, Paris, Alcan, 1901. Leibniz's own inspiration goes back even farther (*Ars Magna*, by Ramon Lull, who died in 1315).
- 109 110. See 'New methods in documentation' (Conference at the ITBTP): 'to symbolize the elementary points of view of the classification of ideas and... to study the grouping of these symbols in order to obtain composite symbols representing the structure of complex concepts' (p. 8), and a little further, the geometrical analogy 'intuition permits the representation in an intellectual space of a logical figure, to n dimensions, a synthesis of the relationships between a group of ideas into the different classes which arrange them naturally according to the various possible individual viewpoints'.
- 109 111. See especially, Robert PAGÈS, *Problèmes de classification culturelle et documentaire* (1955 ed.), op. cit., p. 18-25 'logical classifying structure of Hegelian dialectics', p. 48-9 'introduction to logistic notions', and the entire Chapter 7 'Classification co-ordinates and criteria' (where Pagès criticizes especially the idea of 'arborescent structure' advanced by Cordonnier, p. 91 ff., then refers to 'classification, relationships, propositions', etc.). Pagès' book, it will be recalled, did not appear until 1955, but was already known in 1948 through typewritten copies which had been circulated (in the form of memoranda for the documentation techniques courses at the Union Française des Organismes de Documentation).
- 109 112. See Note 65 above.
- 109 113. H. P. LUHN, 'A new method of recording and searching information', dated 10 September 1951, Appendix I to *The IBM electronic information searching system*, presented in part at the

- Symposium on Machine Techniques . . . MIT, June 1952 (N.Y.: IBM); published also in *American documentation*, vol. 4, no. 1, 1951, p. 14-16.
- 109 114. This is, in fact, the actual sense of the conference convened by the Western Reserve University for the search of 'norms concerning a common language' in September 1959; B. C. Vickery has particularly underlined the point raised here, in his memorandum for the conference, 'A common language for information retrieval', 20 August 1959.
- 109 115. H. P. LUHN, *Scheme of forming words for serial unique identification by searching machines*, 31 March 1953; *Self-demarcating code words* (N.Y.: IBM, 1953; 2nd ed., 1956).
- 109 116. J. W. PERRY, Allen KENT and Madeline M. BERRY, *Machine literature searching*, op. cit., p. 76-79.
- 110 117. PERRY, KENT and BERRY, op. cit.: This paragraph on 'analytic and synthetic relationships' provides a reference to Henry MARGENAU's *The nature of physical reality*, p. 54-122, New York, McGraw-Hill, 1950, as establishing the nature of concepts as 'artifacts', or artificial creations. Perry and his collaborators thereby seem to be connected with a school of thought once made famous by Mach and Duhem, and which one might have believed somewhat outstripped by now. (In the 'acknowledgements' in *Tools for machine literature searching*, p. XI, New York, Interscience Publishers, 1958, Perry and Kent write that 'from the point of view of basic philosophy', (they) 'owe most to the writings of Henry Margenau'.)
- 110 118. ANDREWS and NEWMAN write in report No. 1 of the US Patent Office, *Research and development reports*, p. 12, 15 May 1956, that the distinctions made by Perry-Kent-Berry between their two types of relationships 'do not seem to be usable for the solution of the problem of the Patent Office'.
- 111 119. PERRY, KENT and BERRY, op. cit., Chapter XIII, 'Encoded abstracts', p. 100-8; see, especially, p. 105. In *Tools for machine literature searching* (op. cit.), the punctuation symbols have been changed (see p. 77-8 and p. 146-9). ¶ now indicates the beginning and end of a paragraph, & the beginning and the end of a phrase, and the comma the beginning and the end of a 'phrase element' (subphrase), the asterisk * stands for namely, in so far as, which the / stood for in 1956.
- 111 120. *Ibid.*, p. 92-3 for geographical divisions, p. 86 for divisions into decimal notation.
- 111 121. J. W. PERRY and Allen KENT, *Tools for machine literature searching*, op. cit., with the semantic code dictionary under the general editorship of John L. Melton, of which there is a review by B. C. VICKERY in *American Documentation*, vol. 10, 1959, p. 234-41. In the meantime, *Documentation and information retrieval*, Cleveland, Western Reserve University, 1957, had been published, containing, among others a chapter entitled,

- 'A mathematical model system' for classification systems (p. 4-25) and one on 'Correlation of methods and systems', in which the authors compared various classification systems (the UDC in particular, but not the CC) or 'characterizing systems'; the paragraph relating to 'aspect cards' (visual superimposed cards) indicates wrongly that the latter do not permit the expression of syntactical relationships (p. 127); there is a glossary prepared by REES and KENT, p. 136-50; where the following definitions, among others, are to be found: 'code: any system of symbols in the communication process; particularly a system which achieves abbreviation or some other desirable advantage over common language or numerical expression'; 'analytic relationship: the relationship which exists between concepts (and corresponding terms) by virtue of their definition and inherent scope of meaning'; 'synthetic relationship: a relation existing between concepts which pertains to empirical observation. Such relationships are involved, not in defining concepts or terms, but in reporting the results of observations and experiments.'
- 119 122. The 159th is depolymerization, but polymerization is not included (neither, for that matter, copolymerization). Polymeric is at CYPR.MALC.009, chemical property (attribute), molecule (categoric), No. 9 of this combination of semantic factors. Polyethylene, polystyrene, etc. are found in various decimalized divisions of RASN, plastics.
- 120 123. We shall not attempt to examine here the general question of the treatment of opposition and negation; in this connexion, see the very interesting analysis made by Robert BLANCHÉ, 'Opposition et négation', *Revue philosophique*, no. 2, 1957, 187-216. It sometimes happens that in the PK code one of the opposite notions seems to be lacking. For instance, one finds finite under RAPR.272, but not infinite.
- 121 124. Section E of *Tools for machine searching*, op. cit., p. 379-488.
- 121 125. A series of 213 elemental code-words could certainly have been coded with an average word length of less than four letters, the more so since, in practice, 'self-demarcating' properties of quadrilateral segments have not been used, a period always separating them. Concerning the latter, see the very sound observations made by PAGÈS on the importance of symbolism, and especially of condensed symbols, *Problèmes de classification* . . . op. cit., p. 145-6.
- 122 126. 'Report on the Patent Office', by Wallace Clark and Co., 24 September 1948 (appendix II of the Bush report mentioned in Note 128 below, p. 38). Regarding the history of the classification, see M. F. BAILEY's 'A history of Patent Office classification', *Journal of the Patent Office Society*, vol. 28, 1946, p. 436-507 and p. 537-75.
- 122 127. 'The system of classification now in use has developed in a way that makes the growing up of Topsy seem a model of planned and logical precision.' Warren WEAVER, 'The Patent Office

- problem', *American documentation*, vol. 6, July 1955 (reprinted from *American Patent Law Association Bulletin*, April-May 1955).
- 122 128. *Report (to the Secretary of Commerce) by the Advisory Committee on Application of Machines to Patent Office operations*, Washington, Department of Commerce, 22 December 1954.
- 122 129. Bush report, p. 9-10.
- 122 130. 'A system of retrieval of compounds, compositions, processes, and polymers', prepared by Julius FROME, Jacob LEIBOWITZ and Don D. ANDREWS (Patent Office); Joseph D. GRANDINE, Steven T. POLYAK and Karl G. SIEDSCHLAG, Jr. (duPont de Nemours), 17 November 1958.
- 123 131. LEIBOWITZ, FROME and ANDREWS, 'Variable scope search systems: VS3', *ICSI Proceedings*, vol. 2, p. 1117-42.
- 123 132. LEIBOWITZ, FROME and ANDREWS, 'Variable scope patent searching by an inverted file technique', 17 November 1958. Andrews presented to the ADIA Conference (Frankfurt, June 1959) a report on the same subject 'Automatic categorization of chemical concepts and interrelationships'.
- 123 133. The 'inverted file technique' mentioned here corresponds to what the *IBM information retrieval*, New York, IBM, 1958, refers to as the 'lookup system' (p. 4). Bruno Renard (IBM France), in his typewritten notice 'Recherche de documentation' (1959), used the term 'recherche analytique' (analytical research). If the principle is admitted that in Vickery's 'information matrix' (see his report to the ICSI, *Proceedings*, vol. 2, p. 1278) columns are assigned to 'characteristics' and rows to 'documents', this would amount to 'exploration by columns'—however, this would be somewhat arbitrary (in actual fact, the IBM brochure does assign the columns of the matrix to 'key words' and the rows to 'documents', but Vickery proceeds in an opposite manner, his columns are assigned to 'items' and his rows to 'terms'). Vickery uses the 'term entry systems' denomination (p. 1279). Cordonnier (ITBTP Conference of 11 April 1945, p. 7) used the term 'fiches idéologiques' (ideological cards), which is no longer sufficiently general in scope since computers can be used for 'analytical research'. A standardization of terminology, in this instance, would be absolutely necessary.
- 123 134. The underlying principle of this method of retrieval in several stages, from the generic to the specific, had been outlined by J. J. NOLAN in his report to the American Chemical Society, 15 April 1958 (*American Documentation*, vol. 10, no. 1, January 1959, p. 27-35); the text was slightly modified in the brochure, *Principles of information storage and retrieval using a large scale random access memory*, New York, IBM, 17 November 1958, p. 12-14. The Patent Office report No. 14 (p. 6) adds that the retrieval system outlined acknowledges also 'the role or the function performed by each compound' and that retrieval may also be

- conducted on this basis, as well as on that of identification of the compound, but gives no further details on the subject. A general outline of the so-called 'Haystaq' retrieval system, earlier described in reports Nos. 5 and 8, was presented to the ICSI by Herbert R. KROLLER, Ethel MARDEN and Harold PFEFFER, 'The Haystaq system: past, present, and future', *Proceedings*, vol. 2, p. 1143-79. It discusses, among others, the method adopted of topological representation of molecular structures, which is an adaptation of the Ray and Kirsch method (see below, Note 176).
- 123 135. Simon M. NEWMAN, Reports nos. 1, 4, and 12, all entitled 'Storage and retrieval of contents of technical literature: non-chemical information'. The 'Preliminary Report' (no. 1, 15 May 1956) is signed by DON D. ANDREWS and S. M. NEWMAN; the 'First' and the 'Second Supplementary Report', June 1957 and 17 November 1958, are signed by the latter only. Report no. 9, 'Linguistic problems in mechanization of patent searching, is a general outline; 'Linguistics and information retrieval: toward a solution of the Patent Office problem', *Monograph series in linguistics and language studies*, no. 10 p. 103-11, (Washington, Georgetown University Press, 1957), also reprinted in the *Journal of the Patent Office Society*, 39 (10), October 1957, p. 720-9, summary of Reports nos. 1 and 4.
- 124 136. Reports no. 1, p. 3; no. 9, p. 5; no. 12, p. 5.
- 125 137. Many of the roots published relate to concepts of space. The 'Newman language' is even more 'specific' than ordinary English, for instance, it includes not only triang (triangle), but also isotrix (isosceles triangle), rtriang (right-angle triangle) and sphtriang (spherical triangle).
- 126 138. An interrelational DRIVER (DRIVEN) concept is mentioned in no. 4, p. 5: to be added to table 16 (no. 1, p. 11)—derived from a passage where the authors discussed the difference between relationships having a 'dominant/dominated' character and those which are 'equi-relative'. This DRIVER (DRIVEN) concept disappeared from the no. 12 list, perhaps because the question was never completely elucidated. It would seem that a study of grammatical oppositions in natural languages, like that of oppositions in logic, could clarify the question. Here, we are dealing with the old general distinction active voice/passive voice. We cannot emphasize too much the necessity, in studies of this type, of having teams which—as in operational research—group together specialists of various education; each should include, in addition to the technicians, a mathematician, a philosopher, and a linguist, at the very least. Another aspect of the oppositions is envisaged by Newman under the name of 'dual-aspect roots' (see no. 4, p. 9-11 and p. 15); in connexion with the qualifiers, he also examines the inapplicability of the 'law of the third excluded' (no. 4, p. 15, 2nd column).
- 127 139. 'Analysis of prepositionals for interrelational concepts',

- 15 July 1959. The subtitle indicates that this is a 'preliminary study'.
- 128 140. Let us mention as a reminder that a method of 'serial numeration' designed to complete the interfixes, proposed in report no. 1, p. 9, was later abandoned (see no. 4, p. 11), the retrieval difficulties which it was destined to prevent having been recognized as avoidable by the use of modulants =MCI and =SW.
- 128 141. P. GUIRAUD, *La grammaire*, p. 18, Paris, Presses Universitaires de France, 1958.
- 129 142. It is simply stated therein that 'in the course of the present study, we arrived slowly at the conclusion that we could make use neither of the position of words in a phrase, nor of the grammatical construction of the phrase in order to solve the problem [that of information retrieval at the Patent Office]'.
- 129 143. Walter von WARTBURG, *Linguistic problems and methods*, p. 50-63 (translated by Maillard), Paris, Presses Universitaires de France. He writes that the 'inflected languages have been able to create . . . elements capable of uniting with the word bearing the main meaning; they have also overcome the heaviness of pure juxtaposition and have created a light and flexible tool of the human mind' (p. 49).
- 129 144. In the Leroy-Braffort code, a single code-word (reduced eventually to a monolithic symbol) is used to express a relationship, the direction of the relationship being indicated by the figures 1 and 2 applied respectively to the first and second terms to be joined. These figures perform the role of expression of the order of words: it is known that it is the use of the latter which permits a language, such as Chinese (and to a lesser degree, English and French), to dispense very frequently with inflexions. Such a solution seems more 'elegant' (in the sense of simple and economical which is attributed to this term when it is used in the appreciation of mathematical reasoning) than the apposition of terms, identical or different according to whether they refer to a symmetrical or an asymmetrical relationship to each of the linked concepts, which is the solution of Newman and of the somewhat similar Perry-Kent method.
- 129 145. Charles G. SMITH, 'Descriptive documentation', *ICSI Proceedings*, op. cit., p. 1097-1116.
- 129 146. *Ibid.*, p. 1099; vol. 2, p. 1097-1116. To conceive a 'term' as exclusive of the expression of a relationship is certain to restrict too greatly the use of the word; to conceive of a relationship as a simple empty 'abstract form', on the other hand, is contrary to the reality of facts. But it is seen that it is possible, starting from a false (or perhaps badly expressed) premise to attain good results.
- 129 147. *Ibid.*, p. 1100.
- 130 148. *Ibid.*, 1103. This is somewhat analogous to Cordonnier's

- idea of 'compound symbols' (see above p. 109, Note 110); neither one nor the other seem to doubt the scale of the undertaking and its difficulties, as soon as one leaves the fields (still very restricted) of science already 'formalized'. See in this connexion B. DUNHAM's very interesting article, 'The formalization of scientific language, part I: The work of Woodger and Hull', *IBM journal of research and development*, vol. 1, October 1957, p. 341-8.
- 130 149. *ICSI Proceedings*, op. cit., p. 1098.
- 130 150. *Ibid.*, p. 1105. He gives there a much clearer definition of 'state' than Newman (which was inspired by an earlier note by Smith, as indicated in note 13, of report no. 12).
- 130 151. *Ibid.*, p. 1106 (an interesting discussion of the various aspects under which a machine can be considered), then p. 1107 for an analysis of the example.
- 130 152. *Ibid.*, p. 286-9. See in particular, p. 1113, the 'basic analytical diagram'. It is suggested that this analysis be compared to that of Koelewijn for carburettors (see above, p. 90); they have many points in common, although Smith and Koelewijn, it would seem, had worked independently of each other.
- 130 153. 'Interrelationships as a basis for information retrieval', International Conference for Standards on a Common Language . . ., in: Allen KENT (ed.), *Information retrieval and machine translation*, op. cit., vol. 1, p. 337-44; cf. p. 341. Andrews adds a very true observation to the effect that it is foolhardy to discuss standardization of a 'machine language' or to attempt the latter, so long as a maturely considered and satisfactory solution has not been found to the problem of expression of relationships, for, in the present stage of experimental research on this point, 'each language presently used is satisfactory only in so far as the particular system for which it was specially conceived is concerned'.
- 130 154. Warren WEAVER cited above (Note 127). See p. 131-3 of the text in *American documentation*. It will be recalled that Warren Weaver, vice-president of the Rockefeller Foundation for Natural and Medical Sciences, was, by his famous memorandum of July 1949 (see LOCKE and BOOTH, *Machine translation of languages* (op. cit., p. 15-23), the initiator of research on mechanical translation in the United States. He was a member of the Bush Committee.
- 131 155. Warren Weaver believed a useful schedule of universal classification to be an impossibility (p. 131 of his article). However, if this 'simple logical core' could be found for the entire range of information required at the Patent Office—covering practically the entire field of science and technology—one would be very close to such a schedule. In the same lecture (p. 133 of the text of *American documentation*) there are some very important ideas concerning the fact that an appropriate classification

system to solve a problem such as that faced by the Patent Office should be capable of both 'micro-evolution' and 'macro-evolution'.

- 131 156. H. B. LUHN: see notably *The IBM electronic information searching system*, 1952, already quoted in Note 113, where will be found (in Appendix III) the 'Luko' code formed of pronounceable syllables with 20 consonants and 5 vowels, alternating C-V-C-V; the references quoted above at Note 115 concerning the 'autodemarcatory' codes; *Superimposed coding with the aid of randomizing squares for use in mechanical information searching system*, IBM Engineering Laboratory, Poughkeepsie, 15 June 1956; Luhn also worked on notations for describing multidimensional structures; see his report *A serial notation for describing the topology of multidimensional branched structures*, IBM Research Center, Poughkeepsie, 12 December 1955.
- 131 157. See the rather brief outline given by H. P. LUHN in *The IBM universal card scanner for punched card information searching systems*, p. 11-16, New York, IBM, 17 November 1958.
- 131 158. 'A statistical approach to mechanized literature searching', first published as *Research paper RC-3 IBM* (Poughkeepsie, 30 January 1957), then in the *IBM journal of research and development*, vol. 1, no. 4, October 1957, p. 309-17; 'A business intelligence system', published as a report to the *Conference on Communication of Scientific Information*, San Jose, May 1958; then in the *IBM journal of research and development*, vol. 2, no. 4, October 1958, p. 314-9.
- 132 159. Automatic index preparation by the 'key-words-in-context' method, according to titles of articles; see the report cited above in Note 157, p. 22-4; the bibliography of Peter James cited in Note 89 above was obtained in this manner; see also H. P. LUHN, *Auto-encoding of documents for information systems*, IBM Monograph, 1958, and the appendix of *Information retrieval through row-by-row scanning on the IBM 101 electronic statistical machine (row-by-row scanning attachment)*, p. 4-6, New York, IBM, 17 November 1958, which describes an experiment performed by S. Stiassny. Also: P. B. BAXENDALE, 'Machine-made index for technical literature, an experiment', *IBM Journal of research and development*, vol. 2, no. 4, October 1958, p. 354-61.
- 132 160. 'The automatic creation of literature abstracts', *IBM journal of research and development*, vol. 2, no. 2, April 1958, p. 159-65; *An experiment in auto-abstracting*, New York, IBM Research Center, 1958 (provides auto-analyses of reports in area 5 to the ICSI); T. R. SAVAGE, *The preparation of auto-abstracts on the IBM 704 data processing system* New York, IBM, 17 November 1958.
- 132 161. See Yehoshua BAR-HILLEL, 'The mechanization of literature searching', paper 4-8 for the Symposium on the Mechanization of Thought Processes, Teddington, November 1958; we would not be quite as pessimistic as Bar-Hillel, for experience has

shown that machines for the treatment of information have performed much work previously judged forever beyond their capabilities, owing to too great a complexity. But it is certain that, in order to obtain 'good' auto-analyses, a machine capable of learning would be necessary, which may not be in the too distant future, according to the research now being performed by Mary Stevens and the Manchester group (see p. 139).

- 132 162. H. P. LUHN, *Potentialities of auto-encoding of scientific literature*, p. 5 and p. 12-13, New York, IBM Research Center, 15 May 1959 (Research report RC-101).
- 132 163. H. P. LUHN, *Row-by-row scanning systems for IBM punched cards as applied to information retrieval problems*, p. 8-9 and p. 24, New York, IBM Research Center, 8 May 1959 (Research report RC-100).
- 132 164. The last point seems much more important than he believes. See also F. E. FIRTH, *An experiment in literature searching with the IBM 305 RAMAC*, San Jose, IBM, 17 November 1958.
- 132 165. An experiment was presented on the occasion of the ICSI at the New York IBM. T. T. TANIMOTO's publication, *An elementary mathematical theory of classification and prediction*, New York, IBM, 17 November 1958, unfortunately does not furnish any details. See the short notes on its application to the classification of plants at the New York Botanical Garden under David ROGERS, and later Wm. C. STEEVE in *Science information news*, vol. 1 (6), December 1959-January 1960, p. 5, and *Scientific information notes*, vol. 3 (1), February-March 1961, p. 11.
- 133 166. See the publications of Documentation, Inc., 2521 Connecticut Avenue, Washington, DC; the most recent being the *Operating manual for the Uniterm system of indexing*, 1958, also the four volumes of *Studies in coordinate indexing*, 1953-57, as well as numerous reports under contract NAONR-1305 (00), of which a list will be found in Bourne's bibliography, suppl. 1, p. 15. See also our Unesco report of 1955, Unesco document 320/5601, p. 15-16, and the supplement 320/5 X 30, p. 1. Information for Industry, Inc., has published Uniterm indexes for American patents in fields of chemistry and electronics. The lack of logical precision of certain Documentation, Inc. publications, has provoked criticism by Y. BAR-HILLEL in his interesting article 'A logician's reaction to recent theorizing on information search systems', *American documentation*, April, 1957. On systems of 'control of the vocabulary', see for example Eugene WALL (E. I. duPont de Nemours), 'Use of concept coordination in the duPont Engineering Dept.', report to the ASTIA Conference in Washington, February 1957; *Multiple aspect searching for information retrieval* (use of an 'associative matrix' prepared on IBM cards to furnish a list of frequency of all associations of terms used with approximately 1,800 of the 2,667 Uniterms employed in the system). Eugene Wall established for the duPont Engineering Department a technical thesaurus; he also uses 11

role indicators (see fig. 13 of the report by J. C. COSTELLO, Jr., and Eugene WALL, 'Recent improvements in techniques for storing and retrieving information', to the Society for Advancement of Management, Technical Session, Wilmington, 13 January 1959). See also the contribution of Wall to the inquiry of The Committee on Government Operations of the US Senate, Document no. 113, 86th Congress, 2nd session: *Documentation, indexing and retrieval of scientific information*, p. 175-203 Washington, GPO, 1960; and *Non-conventional technical information systems*, p. 12, Washington, National Science Foundation, 1958. The latter publication (p. 27) describes a similar development in its purpose, although using a different principle, at the National Lead Company, Titanium Division, where a 'subject authority list' is used (official list of Uniterms used) established according to a classification 'based on the principles of the Colon Classification'; mention is made therein of a report on the subject by John WADINGTON, 'Unit concept coordinate indexing', presented to the Division of Chemical Literature of the American Chemical Society in April 1957; we have not seen the latter. The thesaurus method has also been adopted by ASTIA (Armed Services Technical Information Agency); see the brief note on this subject in *Current research and development in scientific documentation (CRDSD)*, no. 7, p. 17-8. Herner & Co. are making a study on the comparative efficiency of indexing with Uniterms and of what they call 'non-manipulative correlative indexing' (*CRDSD*, 7, p. 28-9), and Documentation, Inc. has been commissioned by the National Science Foundation to make a study on the state of the art of 'co-ordinate indexing' (*ibid.*, p. 24).

- 133 167. See particularly his duplicated note 'Description of work done in New Jersey [at the New Jersey Bar Association] on the application of punched cards in the field of law', 9 September 1954; 'Searching legal literature—an appraisal of new methods', *Law Library Journal*, vol. 46, no. 2, May 1953, p. 110-9.
- 133 168. 'Application of punched cards to geologic data concerning uranium deposits in sandstone', US Geological Survey, Denver, 10 August 1956. Published in *Economic Geology*, p. 180-91.
- 133 169. To the article which we cited in the Unesco document 320/5X30, p. 2 and 3, can be added the *Guide to instrumentation literature*, Washington, US Bureau of Standards, 1955, and two publications by Wildhack and Stern (see Peter James' bibliography, p. 21 (2nd ed.)).
- 133 170. Y. S. TOULOUKIAN, C. H. STEVENS, R. H. RODINE, T. WING, and D. W. SMITH, *Systems and procedures developed for the search, coding, and mechanized processing of bibliographic information on thermophysical properties*, p. 4-11. Lafayette, Indiana, Purdue University, School of Mechanical Engineering, Thermophysical Properties Research Center, 1 July 1958. The other categories, apart from those relating to substances, are: property, physical

state, type of study, and strictly speaking bibliographical categories. See also T. WING and Y. S. TOULOUKIAN, 'Substance classification developed for mechanized literature search by the Thermophysical Properties Research Center', *TPRC Report 3*, April 1958.

- 134 171. Developed by Gilbert L. Peakes; summary and bibliographical description in *Non-conventional technical information systems in use*, p. 4-5 (this National Science Foundation publication, cited earlier in Note 166, will hereafter be abbreviated to *NCTI*).
- 134 172. *NCTI*, p. 20-1.
- 134 173. *Ibid.*, p. 35-6. As categories, this company uses, for example, operations, processes, products, equipment, physical properties.
- 134 174. *Ibid.*, p. 18-9; see Fred R. WHALEY, 'Operational experience with Linde's indexing and retrieval system', In: *IBM General Information Manual, Information Retrieval Systems Conference*, 21-23 September 1959. There are 5 role indicators: 2 for absence of the object or of the effect, 3 or 5 to indicate that the term appears as an object of certain verbs, 4 for the agents of an action; in all other cases, the indicator is 1. This elementary syntax bears more resemblance to that of Gardin than to that of Perry-Kent. We shall mention here that Whaley has presented to the ICSI an interesting report on the frequency of the different logical operations implied by the questions posed, 'Retrieval questions from the use of Linde's indexing and retrieval system'.
- 134 175. *NCTI*, p. 37-8 and the report presented to the ACS on 12 September 1955, by W. S. JONES and P. H. BUTTERFIELD, *A technical information service using punched cards for indexing and retrieval*, p. 6, multilithed. The role indicators are: reagents, products, catalysers, chemical agents, construction materials, physical agents; two others indicate the presence in the document of analytic methods or physical properties; a 'negative' code was added subsequently. In *CRDSD*, no. 4, p. 13, April 1959, it was mentioned that W. T. Knox, at the Esso Research and Engineering Co., anticipates changing from a code using simple descriptors to a code expressing their relationships, but he has furnished no details concerning the method which will be followed for the expression of relationships. In *CRDSD*, no. 5, p. 17, it is stated that 'the introduction of liaisons between the descriptors, and of role indicators . . . combined with a more refined method of coding for punched cards, has made it possible to use a simpler machine' (IBM 101, instead of IBM 305 or 704 as previously proposed). *CRDSD*, no. 7, p. 26, mentions a report, which we have not yet seen, by G. Jahoda, M. D. Schoengold and T. J. Devlin, 'A machine-based index to internal research and engineering reports', given in September 1960 to the Division of Chemical Literature of the American Chemical Society.
- 134 176. L. C. RAY and R. A. KIRSCH, 'Finding chemical records by

digital computers', *Science*, vol. 126, 1957, p. 814-9; as we have seen, this system has been adapted to the US Patent Office (see above, Note 134). The Dow method: T. R. NORTON and A. OPLER, *A manual for coding organic compounds for use with a mechanized searching system*, Pittsburgh, Dow Chemical Co., Western Division, Research Dept., 27 May 1953; rev. 15 March 1956; A. OPLER and T. R. NORTON, *A manual for programming computers for use with a mechanized system for searching organic compounds*, Pittsburgh, Dow Chemical Co., Western Division, Research Dept., 25 April 1956; A. OPLER and T. R. NORTON, 'New speed to structural searches', *Chemical & engineering news*, vol. 34, 4 June 1956, p. 2812-16; A. OPLER, 'Dow refines structural searching', *Proceedings of the Western Joint Computer Conference*, 1956 (AIIEE special publication T-85). The method of the Monsanto Chemical Co., Organic Chemicals Division, St. Louis: see *NCTI*, p. 24, and W. H. WALDO and M. DE BACKER, 'Printing chemical structures electronically: encoded compounds searched generically with IBM-702', *ICSI Proceedings*, vol. 1, p. 711-30; R. S. GORDON, J. O. PORTER, W. H. WALDO, 'Routine report writing by computers', *American documentation*, vol. 9, no. 1, 1958, p. 28-31. Although the following is a non-topological system, one may mention the analysis of chemical structures performed at the Cancer Chemotherapy National Service Center of the National Institute of Health, at Bethesda, Maryland; this method, devised by Dean F. Gamble, divides the compounds into four large divisions. It was first used with Uniterms, and later transposed on visual superimposition cards (see *NCTI*, p. 25-6).

- 134 177. L. D. FINDLEY, C. C. BOLZE, R. A. CARPENTER, *A card controlled routine for searching chemical compound data with an IBM 704*, Kansas City, Midwest Research Institute, 17 November 1958.
- 134 178. 'Contributions to the theory of automatic information retrieval', multilithed report, later published in: G. L. PEAKES, A. KENT, and J. W. PERRY (eds.), *Progress report in chemical literature retrieval*, New York, Interscience Publishers, 1957 (*Advances in documentation and library science*, vol. 1). See also A. OPLER and Norma BAIRD, 'Experience in developing information retrieval systems on large electronic computers', *Proceedings*, vol. 1, p. 699-710; this report of a general character makes special reference to experiments in the conversion of codes (which we shall cover later in this work) and a theoretical study, on the basis of symbols composed of letters haphazardly selected, and using research methods of a complicated logical nature. For all these programmes, and on some others, including those of the German firms of Hoechst, and Badische Anilin und Soda-fabrik, see the National Bureau of Standards Report 6865, 'A survey of computer programs for chemical information searching', by Ethel MARDEN and Herbert R. KOLLER, 16 May 1960, which gives an abundant bibliography of 147 items.

- 135 179. 'A computer analysis of the Merck Sharp and Dohme indexing system', duplicated document, undated (1959), prepared by the Remington-Rand Corporation under contract for the Office of Naval Research (Nonr-2297(00) NR 048-116) to appear in *American documentation*.
- 135 180. Cf. *CRDSD*, no. 7, p. 40.
- 135 181. *Ibid.*, p. 42-3. Work done at the System Development Corporation; this body was associated with the researches of Luhn (IBM) on the 'key-words-in-context index' (see Note 159 above) and in fact prepared the first example of such an index: Joan CITRON, Lewis HART and Herbert OHLMAN, *A permutation index to the preprints of the International Conference on Scientific Information*, November 1958.
- 135 182. K. A. KRIEGER, 'A punched-card system for chemical literature', *Journal of chemical education*, 26, 1949, p. 163-6, summarized in the bibliography by Loftus and Kent, p. 142.
- 135 183. The *Preliminary report on research in progress in scientific documentation* of the NATIONAL SCIENCE FOUNDATION (August 1956) referred to a pamphlet of the Avion Division, Alexandria, 'Digitalized logic and its applications', dated September 1955, and also to the report by Miss WILLIAMS, 'Language engineering', at the Conference on Practical Utilization of Recorded Knowledge at Cleveland, January 1956, published later in the collection edited by Jesse H. SHERA, *Documentation in action*, New York, Reinhold, 1956, p. 330-7; the National Bureau of Standards report 6662 refers in its bibliography (p. 12) to a report of ACF Industries, 'Translating from ordinary discourse into formal logic', AFCRC Report TN-56-770, and two other reports by P. M. Williams, made in 1960, are referred to in *CRDSD*, no. 7, p. 33; we have not been able to study any of these reports, except that presented at Cleveland. The Director of the Itek Corporation, J. W. KUIPERS read a paper of a very general character at the ADIA Conference at Frankfurt in June 1959, 'A research program on information searching systems', published on 5 August 1959 as Itek Report P-116; another paper, also not very explicit, is in document no. 113 of the US Senate (86th Congress, 2nd session) *Documentation, indexing, and retrieval of scientific information*, p. 223-30, Washington, Government Printing Office, 1960.
- 135 184. See *CRDSD*, no. 3, p. 24 (where the name MITCHELL is not mentioned: reports by J. P. NASH 'and others' are mentioned, and by C. E. DUNCAN 'and others', under the direction of Louis N. RIDENOUR, LMSD reports 2292, July 1957, and 2366, March 1958), and the bibliography by BOURNE, *ibid.*, Supplement, p. 21. *CRDSD*, no. 6, p. 38, mentions three reports by Mitchell written for contract AF 30(602)-1889, finished in 1959. We have not yet been able to consult any of these studies, which, we are told, have made it possible, by applying the ideas of Lambek and

- of Bar-Hillel, to find an algebraic representation of syntax, 'which covers a large subclass of English sentences'.
- 135 185. Zellig S. HARRIS 'Discourse analysis', *Language*, 28, p. 1-30.
- 135 186. Idem, 'Co-occurrence and transformations in linguistic structure', *Language*, 33, p. 283-340; a detailed analysis in French by Ch. J. BALESTIC, 'La concomitance et les transformations en linguistique structuraliste', internal report of the Commissariat à l'Énergie Atomique (Saclay), DOC/CEN-S/AFD-26 September 1960.
- 135 187. \$421,800 from October 1956 to June 1961 (*Research on mechanical translation*, p. 6, hearings before the Special Investigating Subcommittee of the Committee on Science and Astronautics, US House of Representatives, 86th Congress, 2nd session, May 1960, Washington).
- 136 188. Z. S. HARRIS 'Linguistic transformation for information retrieval', *Int. Conf. Sci. Inf.*, 2, p. 937-50.
- 136 189. See lists in *CRDSD*, no. 2, p. 44; no. 3, p. 38-9; no. 4, p. 33; no. 7, p. 48-9; (the information in nos. 5-6 is repeated in no. 7.)
- 136 190. *CRDSD*, no. 4, p. 33 (April 1959; repeated without alteration in the following numbers).
- 136 191. Noam CHOMSKY, *Syntactic structures*, The Hague, Mouton, 1957. See C. J. BALESTIC, 'Analyse des idées développées par Noam Chomsky dans son livre *Syntactic structures*', internal report of the Commissariat à l'Énergie Atomique (Saclay), CENS-DOC-AFD-2, February 1960.
- 136 192. Victor H. YNGVE, 'The feasibility of machine searching of English texts', *Int. Conf. Sci. Inf.*, 2, p. 975-95; 'In defense of English', report to the Conference at Cleveland of September 1959, published in: Allen KENT (ed.), *Information retrieval and machine translation*, vol. 2, New York, Interscience Publishers, 1960. For the research on mechanical translation carried out by Yngve and his team at the Research Laboratory of Electronics at the Massachusetts Institute of Technology, see *CRDSD*, no. 7, p. 64-5 (the book by Chomsky is mentioned here as reference no. 1).
- 136 193. As witnessed by the translations or abstracts made at Saclay by Balestic, and quoted above in Notes 186 and 191. See also the 'programme of conflicts' of Yves Lecerf at Euratom, above p. 82. In the 'Travaux pratiques de linguistique' which he had prepared for the 'Journée de Linguistique' of the Enseignement Préparatoire aux Techniques de la Documentation Automatique, Lecerf dealt with some linguistic transformations (p. 20-2 of the duplicated text), and he referred here to two of the internal reports (18 and 19) of the 'Transformations and discourse analysis projects' of the University of Pennsylvania, to

which his own method of delimiting the 'frontiers of a field' (of the zone of influence of a word) clearly owes much.

- 136 194. In their article on 'Logical models as a method of scientific research', A. A. ZINOV'EV and I. I. REVZIN (*Voprosy Filosofii*, 14 (1), 1960, p. 82-90, English translation by US Joint Publications Research Service, JPRS 3731, p. 22) write that 'the theory of transformations has great importance for linguistics'. See the article by T. M. NIKOLAEVA, 'What is transformational analysis', *Voprosy Jazykoznanija* (1) 1960, p. 111-5, English translation by US Joint Publications Research Service, 3796, p. 32-41.
- 137 195. HAYS, 'Basic principles and technical variations in sentence structure determination', report to the fourth London Symposium on Information Theory, 1960 (P-1984, p. 1). On the work of the Rand Corporation, see *CRDSD*, no. 7, p. 68-9.
- 137 196. Helen A. PATTERSON, Paul R. ACKLEY, Nancy B. REHMEYER, 'A system for context storage and retrieval of information from the published literature, applicable to both the IBM 101 and electronic computers', *IBM General Information Manual*, 1959 (already quoted in Note 174).
- 137 197. 'Syntactic techniques in information retrieval' (anonymous), National Bureau of Standards report 662, 13 January 1960, p. 1.
- 137 198. Richard S. GLANTZ, 'Further investigation of English syntax with the theory of syntactic types', National Bureau of Standards report 6856, 1 October 1959; R. B. THOMAS, 'The use of SEAC in syntactical analysis' (Monograph series in linguistics and language studies), Georgetown University Press, no. 10 (we have not seen this report, which we know only from the summary given by B. A. and V. A. USPENSKIJ, *Mašinyj perevod i praktičnaja lingvistika* 2 (9) 1959, p. 70-3, English translation by US Joint Publications Research Service 3599, p. 50-2). National Bureau of Standards report 6850, 'A framework for basic research on mechanized information storage, search and selection', 19 May 1960, provides for the use of an 'English recognition grammar' (p. 6).
- 137 199. Don R. SWANSON, 'Searching natural language text by computer', *Science* 132 (3434) 21 October 1960, p. 1099-104. See also the short notes in *CRDSD*, no. 4, p. 26-7 (reproduced without alteration in no. 5, p. 34-5); no. 6, p. 45; and no. 7, p. 39-40. The name of M. E. Maron was originally associated with that of D. R. Swanson; he subsequently disappeared, and was replaced by J. Kuhns and L. C. Ray; finally Paul L. Garvin took charge, with Kuhns, Ray and Swanson under him. A very clear summary is given in the article by Helen L. BROWNSON, 'Research on handling scientific information', *Science* 132 (3444) 30, December 1960, p. 1928-9.
- 137 200. D. E. SWANSON, *ibid.*, p. 1101-2; a specimen of the thesaurus is given.

- 137 201. *Ibid.*, p. 1104, second column. On the question of 'distances' between concepts, see M. DETANT, 'La sphère notionnelle et la notion de distance', *Grisa report*, no. 3, July 1960, p. 7-10 (Euratom document EUR/c/2164/60 f/1), an analytical report which uses the work by CECCATO and LUHN, *Auto-encoding of documents for information retrieval systems* (IBM Monograph, 1958).
- 137 202. F. W. HOUSEHOLDER: research in progress at Indiana University since February 1960, under contract with the Rome Air Development Center, on 'the automatization of general semantics': its purpose is, we read, 'the construction of a regularized artificial language (at first based on English) suitable for storage, mechanical translation or logical manipulation, consisting of a simple structure of phrases, a minimum vocabulary of non-technical words, and a codification procedure for technical words, partly semantic, and partly arbitrary' (*CRDSD*, no. 7, p. 113).
- 137 203. RON MANLY: project of Norair (Division of Northrop Corporation), to construct an 'intelligence language . . . capable of expressing in a comparatively non-ambiguous way practically everything which can be expressed in a natural language, but constructed specially for use on machines'. Manly studies the 'techniques of definition for representing complex concepts in machine language in terms of nuclear concepts which are linguistically not defined' (*CRDSD*, no. 7, p. 121; p. 122 quotes two reports which we have not seen). One could also perhaps include here the research of Marvin S. Cohen in connexion with the 'ACSI-MATIC' system, elaborated by the Radio Corporation of America for the American espionage and counter-espionage service (see *CRDSD*, no. 7, p. 38-9); the text of the documents is 'manually formalized in order to take advantage of a computer's ability to process information'. In general, we have not examined in the present work any 'secret' systems, American or other, for the treatment of information.
- 138 204. This opinion is expressed, in a somewhat naïve form, by Frederick Jonker in his contribution to the inquiry of the US Senate: publication already quoted in Note 183, p. 231-8 (cf. more especially p. 233-5).
- 138 205. 'Les besoins documentaires: cadre dans lequel se situe le contrat avec le GEDSH', *Grisa report*, no. 9, p. 33-53.
- 138 206. BAR-HILLEL, *Some theoretical aspects of the mechanization of literature searching*, Jerusalem Hebrew University, April 1960 (Technical report no. 3); 'The mechanization of literature searching', Paper 4-8 of the Symposium on the Mechanization of Thought Processes, National Physical Laboratory, Teddington, 24-27 November 1958.
- 138 207. See our seventh report to the FID/CA Committee, 'Tendances actuelles en matière de classifications et codifications documentaires', duplicated document, August 1960. (A certain

- number of copies of this document are still available, and we can send them to those of our readers who are interested.)
- 138 208. It is possible that the American linguists, who are the protagonists of this movement, are the victims of their own theories, which aim at excluding semantics from the exclusively linguistic field of investigation. The consideration of the semantic point of view leads very quickly to research in the 'notional fields', and thus to the systematic classification of concepts, the various relationships between these, and general categories, etc. But this is not the place to develop this view.
- 139 209. Mary E. STEVENS: 'A machine model of recall', Unesco, International Conference on the Numerical Treatment of Information, June 1959 (Unesco document NS/ICIP/J.5.4, see especially pages 11-12). It should probably be noted that the work of T. Kilburn, R. L. Grimsdale, and F. H. Summer at the Electrical Engineering Laboratories of the University of Manchester is already proceeding farther along similar lines (see 'Experiments in machine learning and thinking', Unesco report NS/ICIP/5/6/15, a paper presented by the authors to the same conference).
- 139 210. *CRDSD*, no. 4, p. 20-2. M. M. KESSLER has provided us with a copy of his report 'Concerning some problems of intrascience communication' (Lincoln Laboratory, MIT, group report 45-35, 8 December 1958, republished 8 January 1959), which contains supplementary information concerning the research he has undertaken in this field.
- 139 211. The first publication concerning the V. R. ČERENIN coding method is dated 1955: *Nekotorye problemy dokumentacii i mekhanizacija informacionnykh poiskov* [Some problems of documentation and the mechanization of information services], Moskva, Institut Naučnoj Informacii An SSSR. See his report 'The basic types of information tasks and some methods of their solution', *ICSI Proceedings*, vol. 2, 823-53, which is of a rather general character, and especially the paper presented to the International Conference for Standards on a Common Language... Western Reserve University, September 1959, by V. P. ČERENIN, G. A. LAVRENT'EVA, N. V. ZIDKOVA, 'An experimental informational language for mechanized search of scientific-technical literature' (hereafter abbreviated as *EIL*). In: Allen KENT (ed.), *Information retrieval and machine translation*, p. 389-428, vol. 1, New York, Interscience Publishers, 1960.
- 139 212. V. P. ČERENIN and B. M. RAKOV, *Experimental information machine of the Institute of Scientific Information of the USSR Academy of Sciences*, Moscow, 1955 (translation by D. Sobolev of a brochure of the same title in Russian). EIM is somewhat similar to Perry's WRU Selector.
- 139 213. 'The basic modifications of the EIM have been made necessary by the increased complexity of the selection conditions due, in turn, to the transition to the informational language which

includes concrete synthetic links of the single-group type and a double method of reflecting analytic links among the characteristics' (*EIL*, p. 413). Čerenin's definition for the 'synthetic links' (under the name of 'synthetic relations') is found in his report to the ICSI (*ICSI Proceedings*, p. 825): 'connexions which can be established between the characteristics directly on the basis of the content of the e_h [individual informational elements] being indexed or the question'; and that of the 'analytic links' ('analytic relationships') in the same report (p. 828): 'the constant semantic relations established between the characteristics on the force of a broader and older set of information than E' [E = individual e_h information elements in their entirety within which the information retrieval is to be effected]. The 'single-group type' of 'synthetic' relationships is explained in *EIL*, p. 393. This amounts to representing each characteristic of a subject by a group of more general characteristics (of so-called 'second order', i.e., the equivalent of Mooers' 'descriptors') and subsequently to link these second-order characteristics into groups: 'the subject heading is actually defined here not directly by the second-order characteristics, but by their groups'. The 'double mode of reflecting analytic links among the characteristics' is outlined in *EIL*, p. 407-8.

- 140 214. This first stage results in an equivalent of the Taube 'Uni-terms'. *EIL* discusses the Taube method, p. 391-2, by comparing it to that of Mooers, and even to the Colon Classification, and, for that matter, without sufficiently showing the distinctive characteristics of the descriptive features of the Mooers descriptors. See also references to Taube, p. 409 and p. 411 in *EIL*.
- 140 215. *EIL*, p. 416. These categories correspond fairly closely to those of Kent and Perry in their first categorizations of 1952 (*Machine literature searching*, p. 22-3); 'objects', however, linking the two Kent-Perry 'machines' and 'materials' categories, and the category of proper names being added.
- 141 216. *EIL*, p. 403. The 'structural formula' is given at the bottom of p. 403:

$$M\{O[A(ObP)]\}$$

in which M represents method; O operation; A attribute; Ob object, and P process; the part between () represents the 'core' of the subject. Čerenin refers to the research performed by Miss Williams ('Language engineering', quoted in Note 183 above).

- 142 217. A. M. ZUCKERMANN and A. P. TERENTIEV. 'Chemical nomenclature classification', In: Allen KENT (ed.), *Information retrieval and machine translation*, p. 493-501, vol. 1, New York, Interscience Publications, 1960.
- 142 218. See the report by Ida FOREST, 'Automatization of information in the USSR. Some recent research', *Bulletin des Bibliothèques de France*, no. 5, May 1958, p. 407-10. See also L. I. GUTENMAKHER and G. E. VLEDUC, 'The perspectives for the use of machines for the treatment of information in chemistry', report

of Symposium of Eighth Mendeleev Congress on General and Applied Chemistry, 1959 (English translation, US Joint Publications Research Service, R-331-D, NSF 60-72); and G. E. VLEDUC, *Nekotorye voprosy naučnoj informacii v oblasti khimii*, Moscow, 1958 (English translation, US Joint Publications Research Service 3613, 5 August 1960).

- 142 219. HANUS HERZ, 'Die Dokumentation in der Tschechoslovakei', [Documentation in Czechoslovakia], *Nachrichten für Dokumentation*, 9 Jhrg., H. 2, June 1958, p. 69, makes a brief reference to the question of the mechanization of documentation, stating that it is the subject of 'animated discussion'. Generally speaking, the UDC is widely used in the people's democracies. Gabor OROSZ, at the University of Budapest library, has studied from the mathematical standpoint the superimposition codes in punched card machines; see his later article 'Some probability problems concerning the marking of codes into the superimposition field', *Journal of Documentation*, December 1956, and his earlier articles in *Dokumentation*, 1955-56; however, this question relates to 'codification' in the narrow sense of the word (properly speaking, machine language), which we are not considering here. Previously he had written a general article on the problems of information retrieval, *Dokumentation*, vol. 1, no. 9, November 1954, p. 173-8. We have no information whatever concerning what is being done in China, notably at the Scientific Information Institute of the Academia Sinica in Peking.
- 142 220. *Current research and development in scientific documentation*, no. 4, p. 18-9 (reproduced without alteration in no. 5, p. 24-5); *ibid.*, no. 7, p. 33-34. S. PARTHASARATHY presented to the Cleveland Conference of 1959, a report 'Faceted classification as an approach to machine coding'. In: Allen KENT (ed.), *Information retrieval and machine translation*, op. cit., p. 289-94. His report is strictly limited to ideas previously outlined by Ranganathan.
- 142 221. 'Functional operators in engineering language'. In: Allen KENT (ed.), op. cit., p. 295-336.

General categories and the expression of relationships in natural languages, and experiments in international auxiliary languages

- 143 1. ANDREEV, 'The universal code of science and machine languages', p. 1-2, duplicated report to the Cleveland Conference, 1959.
- 143 2. However (what Andreev does not mention), automatic information retrieval is, from another point of view, easier than mechanized translation, since the end language (summarizing

the content of information retrieved) could (at least theoretically) be much simplified in relation to the entrance languages (of the original documents).

- 144 3. Marcel COHEN writes, in the prefatory note to the second edition of *Les langues du monde*, revised under his direction (p. VII, Paris CNRS, 1952), that 'it would be most useful to extract from these descriptions [of languages] catalogues of observable linguistic processes' in *Pour une sociologie du langage*, p. 10, Paris, A. Michel, 1956, he deplores the fact that 'a catalogue of morphological processes in use among known languages has not been established' (and, further, p. 145, 'If descriptive linguistics has become greatly enriched during the 19th and 20th centuries, what is missing, and among urgent tasks to be undertaken, is that of making as complete a catalogue as possible of grammatical methods of expression'). John B. CAROLL in *The study of language*, p. 42, Harvard University Press, 1953, echoes this thought: 'An important step in descriptive linguistics . . . is to delineate the grammatical categories which are mandatory in a given language. . . . This far, linguists have not had occasion to prepare a systematic display of the variety of such grammatical categories to be found in the languages of the world, usually confining their studies to selected languages. This is a task which urgently needs to be done. A considerable portion of the raw data is available, but it needs to be collected and assembled.' It may also be that such a task could only be undertaken with the help of new methods, which could perhaps be somewhat analogous to those applied by J. C. Gardin in other fields of the human sciences: only the use of mechanographic processes would permit the establishment of multiple correlations, without which such 'catalogues' would remain graveyards for factual details, and from which no comprehensive view could be validly extracted.
- 144 4. Furthermore we have been somewhat handicapped from the standpoint of the availability of documents, some of which were not available in Milan—despite the valuable assistance provided by several libraries, especially that of the Università del Sacro Cuore, to which we here express our thanks.
- 144 5. We do not plan to enter here into the philosophical aspect of the 'categories'. It would seem that the early study by P. A. TRENDELENBURG, 'Geschichte der Kategorienlehre', *Historische Beiträge zur Philosophie*, 1, 1846, 23 p., has not been superseded. See Wilhelm WUNDT, 'Zur Geschichte und Theorie der abstrakten Begriffe' [History and theory of abstract concepts], *Kleine Schriften* [Lesser writings], p. 226-58, vol. II, Leipzig, 1911; Franz BRENTANO, *Kategorienlehre* [Study of categories] (ed. B. Kastil), Leipzig, Meiner, 1933; J. E. SALOMAA, 'The category of relation', *Ann. Acad. Sc. Fenn.* (Helsingfors) B XIX, 2, 1929; *Studies in the problems of relation*, Berkeley, University of California, 1930 (Publications in philosophy, XIII). It will be noted that the Russian philosophers are

interested in problems of 'categories': see, for example, V. AFANASIEV's articles on simple and complex categories in *Vopr. Filos.* no. 1, 1956, p. 79-90; those of I. V. BLANBERG on categories of the whole or part, *ibid.*, no. 4, 1957, p. 41-50; of A. POLIKAROV on category of matter, *Deutsche Zeitung f. Philos.*, 4, 1956 p. 539-49; V. V. STOLIANOV on the role and place of philosophical categories in thought, *ibid.*, 5, 1957, p. 672-96. The report by Roger W. BROWN should be noted, 'Language and categories', in: J. S. BRUNNER, J. J. GOODNOW and G. A. AUSTIN, *A study of thinking*, p. 247-312, New York, 1956. On the present structuralist tendencies, see Klaus HANSEN, 'Wege und Ziele des Strukturalismus' [Ways and aims of structuralism], *Zeitschrift für Anglistik und Amerikanistik* 6, 1958, p. 341-81; the paper of Paul DIDERICHSEN to the Eighth International Congress of Linguists (Oslo, 1957), 'The importance of distribution versus other criteria in linguistic analysis', p. 156-82 of the *Proceedings* (notably the criticism by Harris, p. 165-74, and the discussion, p. 194-205); at the same congress the remarks by GARVIN (p. 630-1) on the insufficiency of the criterion of substitution producing 'viable utterance', and those of B. POTTIER on certain difficulties in the application of the distributional criterion and of the 'differential meaning' (p. 590-1). A volume will be available (August 1961) prepared by the Permanent International Committee of Linguists under the direction of Christine MOHRMANN, Alf SOMMERFELT and Joshua WHATMOUGH, *Trends in European and American linguistics, 1930-1960*, Utrecht, Spectrum. We have not seen the publication by Valter TAULI, *The structural tendencies of languages* Helsinki, 1958 (Suomalaisen Tiedeakatemia Tsimituksia, B 115, 1).

- 145 6. In particular the great work: Wilhelm von HUMBOLDT, *Über die Kawi-Sprache auf der Insel Java, nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts* [On the Kawi language on the island of Java, with an introduction on the variety of human language structure and its effect on the intellectual development of the human race], Berlin, Druckerei der Königl. Akad. d. Wissensch., 1836-39. Cf. A. LEITZMANN (ed.), *Werke*, vol. VI, Berlin, 1903; H. STEINTHAL (ed.), *Die Sprachphilosophischen Werke* [Works on the philosophy of language], Berlin, 1884; and see Herman STEINTHAL, *Die Sprachwissenschaft Wilh. v. Humboldt's und die Hegelsche Philosophie*, Berlin, Dümmler, 1848; Daniel G. BRINTON, *The philosophic grammar of American languages as set forth by Wilhelm von Humboldt, with the translation of an unpublished memoir by him on the American verb*, Philadelphia, 1885; Ernst CASSIRER, *The philosophy of symbolic form*, p. 155-63, vol. 1 (see below Note 10, and numerous other references in the index); Lothar KELKEL, 'Reflexions on Wilhelm von Humboldt's philosophy of language', *Études philosophiques*, n.s., 13, October-December 1958, p. 477-85, L. HJELMSLEV, *Principes de grammaire générale*, p. 217, Copenhagen, Host, 1928, relates very closely

- his study of 'concrete grammatical systems' to the 'Innere Sprachform' of Humboldt; similarly, Lucien TESNIÈRE, *Elements de syntaxe structurale*, p. 12-13, 1959, places himself under his aegis. See also H. BASILIUS, 'Neo-Humboldtian ethno-linguistics', *Word*, 8, 1952, p. 95-105.
- 145 7. See A. MEILLET's article 'Ce que la linguistique doit aux savants allemands', *Scientia*, 1923, reproduced in his collection *Linguistique historique et linguistique générale*, p. 152-9, vol. II, Paris, Klincksieck (new edition, 1951). The latter contains no mention, however, of the name of G. von der Gabelentz, whose great Chinese grammar has not yet been superseded, and, as Hjelmslev in *Principes . . .*, op. cit., p. 215, points out, it was he who, in 1891, was first to use the word 'system' for language, in his *Sprachwissenschaft* [Linguistics].
- 145 8. A. MARTY, 'Über die Scheidung von grammatischen, logischen und psychologischen Subjekt, resp. Prädikat' [On the separation of grammatical, logical and psychological subjects], *Archiv für systematische Philosophie* 3, 1897, p. 174-90 and 294-333; 'Über das Verhältnis von Grammatik und Logik' [On the relationship between grammar and logic], *Symbolae Pragenses*, 1893, p. 99-126; *Untersuchungen zur Grundlegung der allgemeinen Grammatik und Sprachwissenschaft* [Investigations on the basis of general grammar and linguistics], Halle, 1908; *Psyche und Sprachstruktur* [Psyche and linguistics], Bern, Francke, 1940; etc. See Otto FUNKE, *Innere Sprachform: eine Einführung in A. Marty's Sprachphilosophie* [The inner language: an introduction to A. Marty's language philosophy], Reichenberg, 1924; and *Studien zur Geschichte der Sprachphilosophie* [Studies on the history of the philosophy of language], Bern, 1927.
- 145 9. Wilhelm WUNDT, *Völkerpsychologie: I, Die Sprache* [Ethnic psychology: I, Language], Leipzig, 1900 (3rd edition, 1911-12); *Sprachgeschichte und Sprachpsychologie* [History and psychology of language], Leipzig, Engelmann, 1901. See concerning *Die Sprache* A. MEILLET's severe criticism in *Année sociologique*, 5, p. 595-601 (extracts in M. COHEN's *Pour une sociologie . . .*, op. cit., p. 21-2). It will be noted that the classification of parts of speech found in the introduction to the second volume of *Die Sprache* (substantive, adjective, verb, relationship words) has inspired many subsequent authors: see O. FUNKE, *Seventh International Congress of Linguists, London, 1952, Proceedings*, p. 259-60, London, 1956. Braffort and Leroy have revived it (probably by mere coincidence) as well as G. Patrick MEREDITH, 'Semantic matrices', *ICSI Proceedings*, 1958.
- 145 10. Ernst CASSIRER, *Philosophie der symbolischen Formen: I, Die Sprache* [The philosophy of symbolic forms: I, Language], Berlin, B. Cassirer, 1923 (English translation by Ralph Manheim, New Haven, Yale University Press, 1953). This important book deserves detailed study, which should distinguish, however, any valid analyses still to be found on the expression of space, spatial

relations, time, number, and other general concepts in languages, from the neo-Kantian philosophical theories peculiar to the author. As regards the influences which have been brought to bear upon him (apart from Kant, and also Leibniz), first Humboldt, and later Wundt, should be mentioned. Cassirer had exerted considerable effort in documenting himself on a great number of languages, deriving much benefit (as he, himself, says on p. 72 of the English translation) from Meinhof's advice: he had made use of F. Boas' great work. But too frequently he systematizes on fragile bases, which subsequent research in his main sources have demonstrated as non-valid. This becomes evident, for example, on reading what he has written regarding grammatical gender (p. 295-302) where, especially, everything concerning Bantu multiple asexual genders is completely superseded by modern interpretations (among others L. HOMBURGER, *Les préfixes nominaux . . .*, 1929). See also the two articles written by Cassirer toward the end of his life in the United States, and published after his death: 'L'influence du langage sur le développement de la pensée dans les sciences de la nature', *Journal de psychologie normale et pathologique*, 39, 1946, p. 129-52; 'Structuralism and modern linguistics', *Word*, 1, 1945, p. 99-120; and see C. H. Hamburg, *Symbol and reality; studies in the philosophy of Ernst Cassirer*, The Hague, Nijhoff, 1956.

- 145 11. Mainly in Jost Trier, and later Johann Leo Weisberger; see Stephen ULLMANN, *The principles of semantics*, Glasgow, Jackson, 1951; 2nd ed., 1957, p. 154-68 and 309-13, which provides the bibliography. We shall see further the works of W. von Wartburg.
- 145 12. Field in which the important work of H. GLINZ, *Die innere Form des Deutschen* [The inner form of the German language], Bern, 1952—of which the title indicates that the Humboldtian tradition remains alive—seems somewhat isolated.
- 145 13. MEILLET, *Linguistique historique et linguistique générale*, 1, 7, Paris, Champion, new edition, 1948. As a matter of fact, as indicated by O. Funke, the first 'structuralist' worthy of the name may have been the Frenchman Pierre de La Ramée (Petrus Ramus), with his *Gramère* of 1562, containing a division of the parts of speech based on purely formal criteria.
- 145 14. Meillet himself, in his preface to *Linguistique historique et linguistique générale*, p. VIII, op. cit. (this work will hereafter be abbreviated to *LHLG*), makes a significant reference to Saussure. It is not by chance that, in the title of his manual, the word 'historical' is given first place. See P. GUIRAUD, *La grammaire*, p. 78, Paris, Presses Universitaires de France, 1958: 'Structural grammar has resulted from the teachings of F. de Saussure and his course in general linguistics.' It would not be appropriate to discuss here Meillet's own contribution and what may be deduced from it as regards our own subject, i.e. the study of general

categories. One can find in his work not so much on their definition and their detailed characterization, but rather some basic ideas on the trends and factors of their evolution—ideas which, however, would still have to be subjected to a critical test of their agreement with the main facts outside the Indo-European field (and even within this field with periods older than those which Meillet, at the time of his first research in general linguistics, was able to study). In this connexion, the six reports (dating from 1909 to 1920) forming pages 130-229 of *LHLG*, I, should be noted and particularly, the use of the 'oppositions' formula (under the influence of Saussure?) in the article 'Sur les caractères du verbe' (1920). It should also be observed that apparently Meillet, once he had constructed his theory on the 'universal tendency of language' 'to constitute a word with a constant form representing a general idea' (see 'Le caractère concret du mot', 1922, *LHLG*, II, p. 12-13), no longer wished to (or could) change it, although new facts concerning the pre-inflectional state of the Indo-European (which he mentions himself in his contribution to the *Meinhof Festschrift*, 1928; *LHLG*, II, p. 50-1) implied the need for a revision. Other aspects of Meillet's work have aged: for example, the idea that 'any specific definition in general morphology [a term which, for him, covers the entire grammatical system, see *LHLG*, I, p. 83-4] is outside linguistics' (article of 1928, 'Sur la terminologie de la morphologie générale', *LHLG*, II, 34); he is seen here (at the time of Hjelmslev's *Principes* . . .) to be far removed from the movement which, precisely, was to seek to constitute grammatical categories strictly established on 'formal' bases derived from a study of the structures peculiar to language.

- 145 15. Ferdinand DE SAUSSURE, *Cours de linguistique générale*, 1st ed., 1916; 2nd ed., Paris, Payot, 1922, p. 141-92.
- 146 16. Which can be dissociated from those of Saussure's general theories, for example, concerning 'signified' and 'signifying', or concerning 'language' and the 'spoken word'. See, among others, those relating to the philosophico-social concepts of S. W. DOROZEWSKI, 'Durheim and F. de Saussure', *Journal de psychologie normale et pathologique*, 1933, p. 82-91.
- 146 17. From our present point of view: C. BALLY, *Linguistique générale et linguistique française*, 1932; 2nd ed., Bern, 1944; *Traité de stylistique française*, 1909; 3rd ed., reprint, Geneva and Paris, Klincksieck and George, 1951 (see especially the chapter 'Fondements rationnels de la synonymie', vol. I, p. 140-54, and the appendix, 'Tableau synoptique des termes d'identification et de leurs principaux synonymes', vol. II, p. 223-64).
- 146 18. As indicated by HJELMSLEV, *Prolegomena to a theory of language*, p. 51, note) the method of 'description of categories of expression' has been 'lucidly formulated' by SÉCHEHAYE in his *Programme et méthodes de la linguistique théorique*, Paris, 1908; see also *Essai sur la structure logique de la phrase*, Paris, Champion, 1926; reprinted, 1950.

- 146 19. H. FREI, *La grammaire des fautes*, Paris, Geuthner, 1929, a remarkable study on popular 'advanced French'. See the *Cahiers Ferdinand de Saussure*, a periodical which is the organ of the present Geneva school.
- 146 20. Gustave GUILLAUME, *La pensée et la langue*, Paris, Masson, 1922. BRUNOT reacted with good reason against the traditional grammar with its pseudological categories; he was doubtless under Bally's influence (see, in the *Traité de stylistique*, vol. I, p. 257-8, the projected construction of a syntax 'which would proceed from thought, with a view to studying its linguistic creations', starting from 'the formal aspects of thought, or in other words, form-ideas'—the term 'form-ideas' is found again in Brunot—and seeking the 'grammatical types assumed by these form-ideas in a given language at a given period'). Despite the apparent opposition to Brunot indicated by the sub-title of their *Essai de grammaire de la langue française; des mots à la pensée*, Paris, D'Artrey, 1911-52, this attempt by J. DAMOURETTE and E. PICHON is basically of the same kind, bringing little more to scientific linguistics than a tremendous collection of detailed material. See, concerning Brunot, the judgements expressed or quoted by Gérard ANTOINE, *La coordination en français*, vol. I, p. 105 and p. 130, Paris, D'Artrey, 1959.
- 146 21. 'The idea of language, a sort of intangible location of thoughts acting within it in a systematic manner, has become a kind of idealistic myth whose most developed stage is probably found in the most recent works of G. Guillaume', Marcel COHEN, *Pour une sociologie . . .*, p. 89, op. cit. See, however, G. ANTOINE'S essay justifying the 'sub-linguistic scheme', *La coordination en français*, p. 56, op. cit.
- 146 22. For example in GUILLAUME'S work 'Observation et explication dans la science du langage', *Études philosophiques*, 13, October-December 1958, p. 446-62, where we find phrases of this type (p. 458): 'La dyade livrée par la linguistique cryptologique c'est: mentalisme de signifiante décroché de son physisme de représentation/mentalisme de subsignifiante (mentalisme de soubassement) éclairant par en-dessous le mentalisme de signifiante superposé' [The dyad formed by cryptological linguistics is: mentalism of significance detached from its physism of representation/mentalism of sub-significance (sub-mentalism) illuminating from below the mentalism of superimposed significance], or when Guillaume criticizes (p. 456) 'traditional linguistics' for having had 'an insufficient premonition of what is represented by a transnullity of nullity when the outstripped nullity is that of a *see* in a cryptological syndese relationship to a hypodese . . .' (sic).
- 146 23. 'The human mind', writes Guillaume, circumvents the difficulty which is a result of the fact that 'universal perception cannot be outdistanced, it is a perception which has no "beyond";' in 'opposing the universe to himself under conditions of

antinomic understanding which are called *space* and *time*. More exactly, *universe-space* and *universe-time*. The grammatical distinction of the noun and of the verb is merely *the linguistic expression of these two visions of the universe*. There is some justification therefore in defining the noun as the word whose understanding is completed outside time, in space; and the verb as the word whose meaning is completed in time.' 'Comment se fait un système grammatical', *Conférence de l'Institut de linguistique de l'Université de Paris*, VII p. 56-7, 1939. One would seem here to be going beyond linguistics into mysticism.

- 146 24. In Guillaume's bibliography in Roch VALIN's *Petite introduction à la psychomécanique du langage*, Quebec, Presses Universitaires Laval). See especially, *Le problème de l'article et sa solution dans la langue française*, Paris, Hachette, 1919 and subsequent articles on the same subject in: *Français moderne*, 12, p. 89-107; 13, p. 70-82 and p. 207-29 (1944-45); *Temps et verbe*, Paris, Champion, 1929; 'Théorie des auxiliaires et faits connexes', *Bull. Soc. Ling. Paris*, 39, p. 5-53; *L'architecture du temps dans les langues classiques*, Copenhagen, Munksgaard, 1945; *Époques et niveaux temporels dans le système de la conjugaison française*, Quebec, Presses Universitaires Laval. G. Guillaume's disciples were: R. L. WAGNER, who analysed his works in his *Cours de grammaire et philologie*, Paris, 1953; see also his article 'Coordonnées spatiales et coordonnées temporelles', *R. Ling. rom.*, 12, 1936, p. 144-64; Roch VALIN, *Esquisse d'une théorie des degrés de comparaison*, Quebec, Presses Universitaires Laval; and more recently, Bernard PORTIER, whose very important work on *Les éléments de relation en français et en espagnol*, Paris, Klincksieck, is awaited next summer.
- 146 25. Lucien TESNIÈRE, *Éléments de syntaxe structurale*, Paris, Klincksieck, 1959. As Hjelmslev did, Tesnière assigned a preponderant place to the direction notion; furthermore, affinities will be recognized between his method of representation by stemmas and the techniques of Ceccato (and of Braffort). We merely wish to mention this book of approximately 700 pages, and not to analyse (and even less to criticize) it.
- 147 26. Emile BENVENISTE, 'Le système sublogique des prépositions en latin', *Travaux du Cercle Linguistique de Copenhague*, 5, p. 177-84—an enlightening study on the distinction between *pro* and *prae*; see also 'Tendances récentes en linguistique générale', *Journal de psychologie normale et pathologique*, no. 1-2, 1954, p. 130-45.
- 147 27. A. MARTINET, primarily a phonologist (see his *Economie des changements phonétiques*), was among the first to comment on the 'Prolégomènes à une théorie du langage' of Hjelmslev ('Au sujet des Fondements de la théorie linguistique de Louis Hjelmslev', *Bull. Soc. Ling. Paris*, 42, 1946, p. 19-42); during his long sojourn in the United States he familiarized himself with Ameri-

can structuralism (see 'Structural linguistics', In: KROEBER, *Anthropology today*, p. 574-86, Chicago, 1953). Under his direction, the *Travaux de l'Institut de Linguistique* of the University of Paris (to volume 1 (1956) of which, he contributed a programme article 'Linguistique structurale et grammaire comparée') published in 1957 an important international investigation on the 'Notion of neutralization in morphology and the lexicon'. See also his recent *Éléments de linguistique générale*, Paris, Colin, 1961. We have not mentioned Raoul de La Grasserie, whose numerous (and rather uneven) works should perhaps be recalled, published between 1887 and 1914, and of which a list can be found in Hjelmslev's bibliography, *Principes de grammaire générale*, op. cit. Claude LÉVI-STRAUSS translated in his book *Anthropologie structurale*, p. 37-91, Paris, Plon, 1958, three of his articles published first in English in the United States, followed (p. 93-110) by a reply to a critique by Haudricourt and Granai; suggestive parallels will be found there between linguistic structures and other social structures, especially those of relationship.

- 147 28. H. G. WIWEL, *Synspunkter for dansk sproglære* [Aspects of Danish linguistics], 1901; see HJELMSLEV, *Principes . . .*, p. 109, op. cit.
- 147 29. Otto JESPERSEN, *The philosophy of grammar*, Chapter VII, 'The three ranks', p. 96-107, London, Allen, 1924. See Hjelmslev's *Principes . . .*, p. 161-2, op. cit., which refer to this three ranks theory of Jespersen, of which he considers the logical basis as superfluous. A little earlier he writes: 'La transitivité, la direction, est le principe constituant fondamental de toute organisation grammaticale' [Transitivity, direction, is the fundamental constituent principle of any grammatical structure] (p. 154). Diderichsen mentions, in fact, that 'the traditional concepts of government and concord are the germs of glossematics, and direction is intended to a precise and more convenient reinterpretation of these ancient terms' (*Travaux Cercle ling. Copenh.*, 5, 1949, p. 152). H. J. ULDAL ('On equivalent relations', *ibid.*, p. 71) also mentions the origin in Jespersen of fundamental notions of glossematics on relationships, which he considers to be more general, and their notation by arrows clearer and simpler.
- 147 30. See mainly Viggo BRØNDAL: *Ordklasserne* [The parts of speech], Copenhagen, Gad, 1928 (French translation by Pierre Naert, *Les parties du discours*, Copenhagen, Munksgaard, 1948); *Propositionernes teori* [Theory of prepositions], programme of the University of Copenhagen, 1940 (translation by P. Naert, *Théorie des prépositions*, Copenhagen, Munksgaard, 1950); and the articles or notes assembled in a posthumous handbook *Essais de linguistique générale*, 1943.
- 147 31. *Essais . . .*, op. cit., p. XII.

- 147 32. Especially Marcel COHEN, in his reviews of the *Essais...* (*Bull. Soc. Ling. Paris*, 47, 1951, p. 23 and *Année sociologique*, 2, 1940-48, p. 844), and of the *Théorie des prépositions* (*Bull. Soc. Ling. Paris*, 47, 1951, 23-5, and *Année sociologique*, 1951, p. 493). Similarly, O. FUNKE writes concerning *Les parties du discours* (*Proceedings of the 7th International Congress of Linguists*, p. 260-1) that this is a 'highly ingenious play', of which the principal weakness resides in the 'imperfection of psychological analysis...'.
 147 33. Furthermore, in Brøndal there are highly questionable speculations where he seeks to relate—on the basis of his distinctions between 'abstract' and 'concrete' systems—such or such a grammatical fact to a kind of 'mentality', primitive or otherwise: notably in *Les parties du discours* (p. 166-71), e.g., he writes that the Indo-European languages 'seem to be the only ones to have attained full development both of abstract and concrete classes... In other words, a more complete mentality will be found among the Semites and the Indo-Europeans, as well as a greater diversity of faculties than among other peoples'; in the passage on the 'relationship between language and thought' of the *Théorie des prépositions* (p. 120-1), or, again, in the *Scientia* article of August 1935 'Structure and variability of morphological systems' (*Essais*, p. 23) where, in outlining his personal conception of linguistic oppositions (comprising, apart from a simple duality, a neutral term, a complex undivided term and complex-negative and complex-positive terms), he writes 'it is therefore to be anticipated that the progress of the human mind... will find expression in language by the acquisition of neutral forms and by the loss of complex forms'. Brøndal's works can probably be defined as a bold attempt, but a premature and insufficiently objective one, to introduce into grammar the methods of analysis which had been successful in the field (basically incomparably less complex) of phonology. Nevertheless, indications are to be derived from his research—and verified. It may be noted also that, with formulae such as those of *Les parties du discours* (p. 66-7) (e.g., 'The syntactic function of a word should not... determine its classification'), Brøndal places himself well outside all present thinking, which assigns to the 'syntactic behaviour' of words a preponderant role in their classification into parts of speech.
- 147 34. B. SIERTSEMA, *A study of glossematics*, The Hague, Nijhoff, 1955; the book is rather disappointing.
- 147 35. The basic outline is the *Omkring sprogteoriens grundlaeggelse* (Festschrift udgivet af Københavns Universitet, November 1943), translated into English by F. J. WHITFIELD under the title *Prolegomena to a theory of language*, Baltimore, Waverly Press, 1953 (Memoir 7 of the *International Journal of American Linguistics*): In addition to Martinet's article mentioned above (Note 27), P. L. GARVIN's reports, *Language*, 30, 1954, p. 69-96, may be consulted. An *Outline of glossematics* by L. HJELMSLEV

and H. J. ULDALL, of which a 'Synopsis' had appeared in the form of a preliminary printing in 1936, was published (vol. I, by Uldall) in Copenhagen in 1952. The 'Recherches structurales', *Travaux Cercle ling. Copenh.*, V, 1949, provide HJELMSLEV's bibliography to that date, of which we should at least cite here his first important work, *Principes . . .*, op. cit., the formulations of which, however, were subsequently superseded by his own works: 'Essai de théorie de morphèmes', *Proceedings of the 4th International Congress of Linguists*, 1936, p. 140-51, Copenhagen, Munksgaard, 1938; 'La catégorie des cas', *Acta Jutlandica*, VII, I, 1935 and IX, 2, 1937; 'La notion de rection', *Acta linguistica*, I, 1939, p. 10-23; 'La structure morphologique', *Rapports du 5e Congrès International des Linguistes*, p. 66-93, Brussels, 1939; 'Structural analysis of language', *Studia Linguistica*, I, 1947, 69-78; 'Le verbe et la phrase nominale', *Mélanges Marouzeaus*, 1948, p. 253-81; 'Rôle structural de ordre des mots', *Journal de psychologie normale et pathologique*, January-March 1950, p. 54-8; 'In what measure may the significance of words be considered as forming a structure?' *Reports for the 8th International Congress of Linguists*, Oslo, 1957, II. Within the *Prolegomena to a theory of language* (p. 50-1, note, op. cit.), will be found a list of works on descriptive linguistics based on the glossematic theory, of which one only will be mentioned here, that of Knud TOGEBY, *Structure immanente de la langue française*, Copenhagen, 1951. The *Grammatica estructural*, Madrid, Grédos, 1951, of Emilio Alarcos LLORACH for the Spanish language has been criticized (among others by F. J. WHITFIELD, *Word*, 9, p. 279-80) for too many arbitrary simplifications. See also Jens HOLT's article, 'Rational semantik (pleremic)' [Rational semantics], *Acta Jutlandica*, XVIII, 3, 1946. We have not examined another Danish work of interest here on the same subject, by R. MAGNUSSON, *Studies in the theory of parts of speech*, Copenhagen, Munksgaard, 1954, criticized by F. J. WHITFIELD, *Language*, 31, 1955, p. 245-7, as based too entirely on logical criteria. We have seen, but have not studied, H. SPANG-HANSEN's *Probability and structural classification*, Copenhagen, Rosenkilde.

- 148 36. BRØNDAL, *Les Parties du discours*, p. 165, note 1, op. cit., criticizing it for adhering 'rather to differences of temperament, which are innumerable and vary from person to person . . . than to truly fundamental types of spiritual structure'.
- 148 37. An exception must be made in the case of J. R. FIRTH: *The tongues of men*, London, 1937; *Papers in linguistics*, 1934-1951, London, 1957; 'Structural linguistics', *Trans. Philological Society*, 1955, p. 83-103; and also in the case of M. A. K. Halliday, a member of the Cambridge Group. We have been unable to document ourselves on the works of the Scottish structuralist John C. Catford (see Fourquet's preface to Tesnière's book mentioned earlier, Note 25). An entire English school, following Wittgenstein, is engaged in a study of relationships between

- language and philosophy, but this does not concern us directly here.
- 148 38. John B. CARROLL, *The study of language: a survey of linguistics and related disciplines in America*, Cambridge, Harvard University Press, 1953.
- 148 39. Charles F. HOCKETT, *A course in modern linguistics*, New York, Macmillan, 1958.
- 148 40. Franz BOAS, *Handbook of American Indian languages*, Washington, Government Printing Office; 1911-22, 2 vol.; part 3, New York, Augustin, 1933-38). It is also F. Boas who founded in 1917 the *International Journal of American Linguistics*.
- 148 41. Essentially due to Leonard BLOOMFIELD's *Language*, New York, Holt, 1933, which was preceded by *Introduction to the study of language*, New York, Holt, 1914; the former is presented by the author as a revision of the latter; however, in 1914 Bloomfield based his theory on Wundt while in 1933 he affirms the autonomy of linguistics in relation to psychology (p. VII-VIII of the preface, English edition of 1935). It is certain that chapters 10-16 (p. 158-280, same edition) of this book, devoted to grammatical forms, were at that time far in advance of anything found elsewhere—including Hjelmslev's first book. More so than any other before him, Bloomfield insists on the necessity of 'identification in formal terms' without recourse to 'signification', see for example the passage on page 266: 'Class meanings, like all other meanings, elude the linguist's power of definition, and in general do not coincide with the meanings of strictly-defined technical terms. To accept definitions of meaning, which at best are makeshifts, in place of an identification in formal terms, is to abandon scientific discourse'; or the one on page 185 concerning 'taxemes of selection', where he defines the English noun and verbal expressions in a purely formal manner ('The positions in which a form can appear are its *functions*, or, collectively, its *function*. All the forms which can fill a given position thereby constitute a *form-class*'); or that on p. 271, where he writes that it is necessary to 'determine the English parts of speech not by their correspondence with different aspects of the practical world, but merely by their functions in English syntax'.
- 148 42. Edward SAPIR, *Language*, New York, Harcourt Brace, 1921. Perhaps not as strict as similar chapters in Bloomfield's book, and less exclusively 'formalist', Sapir's outline which he symptomatically entitled in 1921 'Form in language', is not basically different—we shall refer again to Sapir, on the subject of typology or international auxiliary languages. In the work edited by David G. MANDELBAUM, *Selected writings of Edward Sapir in language, culture and personality*, Berkeley, University of California Press, 1949, one can read the admirable article published in H. L. MENCKEN's *American Mercury*, p. 150-9, vol. 1, 1924, 'The

- grammarians and his language', in which Sapir studies very closely the relationships between 'form and 'function'.
- 148 43. Harry HOIJER, under whose direction the handbook *Linguistic structures of native America*, New York, Viking Fund, 1946, was published; it contains formal descriptions of 13 American-Indian languages, of which the plan had, in fact, been established by Sapir.
- 148 44. Benjamin Lee WHORF, *Language, thought, and reality*, New York, Wiley, 1956 (collection of studies written from 1929 to 1942); see also his two monographs on Hopi and Aztec in Hoijer's handbook (p. 158-83 and p. 367-97), which constitute undoubtedly its most remarkable section; Whorf is particularly known for his theory on the relationships between language, on the one hand, and science or philosophy on the other, which we shall encounter again later on. But less attention has perhaps, wrongly, been paid to some of his ideas on the description of language (distinction between 'overt' and 'covert' categories, 'selective' and 'modulus'; notion of 'cryptotype'); see especially 'Grammatical categories', written in 1937 and posthumously published in 1945 (p. 87-101, of the 1956 collection).
- 148 45. Joseph H. GREENBERG, *Essays in linguistics*, University of Chicago Press, 1957 (published also by the Wenner-Gren Foundation as Viking Fund Publications in Anthropology No. 24), notably chapter 8, 'Order of affixing: a study in general linguistics', where he develops a Sapir idea (p. 86-94). Furthermore, Greenberg attempted (chapters 1 and 2, 'Language as a sign system' and 'The definition of linguistic units', p. 1-34) to introduce the study of language into that much vaster study of the systems of signs with the aid of logico-mathematical methods, and has examined, with frequently new and penetrating viewpoints, certain questions of the evolution of language; we shall return to this subject.
- 148 46. Zellig S. HARRIS, *Methods in structural analysis*, University of Chicago Press, 1951; Yehoshua Bar-Hillel criticized very severely (but in rather general terms) Harris's methods of 'analysis by transformations' in his *Report on the state of machine translation in the United States and Great Britain*, p. 16-8, Jerusalem, 1959, blaming them, in particular, for using notions of 'transformation' and of 'core' in an ill-defined sense. His 1951 book has also given rise to some criticism: among others, J. R. FIRTH, at the Seventh Congress of Linguists (*Proceedings*, p. 182-3) blames him for using the word 'meaning' 'in an extraordinary way', on p. 166-71 and p. 195 of *Methods in structural analysis*. Marcel COHEN was particularly severe in his article 'Linguistique moderne et idéalisme' written in response to questionnaires distributed by Soviet linguists, published first in Russian in 1958, and subsequently in *Recherches internationales* 7, May-June 1958 (see p. 70-1 where he speaks of 'linguistic quarterings', 'sophism of identification by "distribution" only

- and attempts to draw nearer to the theories of Marr, which does not seem to be justified). Without going as far as adopting such a purely negative attitude, it would surely be permitted to support the more moderate reservations expressed by J. B. CARROLL (op. cit., p. 31-2, see Note 38 above) and by Stephen ULLMANN (op. cit., p. 317-21, see Note 11 above). A detailed discussion cannot be undertaken here, it would be necessary, to examine especially: the 'appendix to 12.41: The criterion of meaning', (p. 186-95 of *Methods in structural analysis*); the 'vague meaning characteristics' of the 'classes of classes of morphemes' (note 21, p. 252); the 'considerations of meaning' which 'cannot be used other than historically', 'as a source of hints' (note 6, p. 365); it would also be necessary to evaluate the justification for formulae such as that on the 'deductive system' which 'would permit each one to synthesize or predict subject matter contained in the language' analysed (p. 372-3); and to ask whether language descriptions given in the form of the 'lists' enumerated (p. 376-8) could really serve, in themselves, the ambitious 'correlations' envisaged (p. 374-5), without any other less mechanical elaboration.
- 148 47. Charles C. FRIES, *The structure of English*, New York, Harcourt, Brace, 1952.
- 148 48. Noam CHOMSKY, *Syntactic structures* ('s-Gravenhage, Mouton, 1957); 'Three models for the description of language', *IRE Transactions on information theory*, IT-2, 1956, no. 3, p. 113-24; *Semantic considerations in grammar*, Washington, Georgetown University, November 1955 (Institute of languages and linguistics, Monograph No. 8). Chomsky was a student under Harris, but subsequently his research developed independently—and extends further. In the same sense as Chomsky: J. LAMBEK, 'On the mathematics of sentence structure', *Minutes Proc. R. Soc. Canada*, 1956, 50, appendix C, 10; 'The mathematics of sentence structure', *American mathematical monthly*, 65, 1958, 154-70.
- 149 49. Roman JAKOBSON: 'Zur Struktur des russischen Verbums', [On the structure of the Russian verbs], *Charisteria Gvilelmo Mathesio Oblata*, Prague, 1939, p. 72-84; 'Beitrag zur allgemeinen Kasuslehre (Gesamtbedeutung der russischen Kasus)' [Contribution to the general theory of Russian cases], *Travaux Cercle ling. Prague*, 6, 1936, p. 240-88; 'The phonemic and grammatical aspects of language in their interrelations', *Actes 6e Congrès Int. Ling.*, 1948, 5-18 and p. 601, Paris, Klincksieck, 1949; *Shifters, verbal categories, and the Russian verb*, Harvard University Press, 1957. See, in the book in tribute *For Roman Jakobson* ('s-Gravenhage, Mouton), the contributions of Carl L. EBELING, 'On the verbal predicate in Russian' (p. 83-90); A. W. DE GROOT, 'Classification of cases and uses of cases' [re: Latin] (p. 187-94); E. M. UHLENBECK, 'Verb structure in Japanese', (p. 567-73); Hans VOGT, 'Remarques sur la structure formelle du verbe basque' (p. 600-4).

- 149 50. Joshua WHATMOUGH, *Language: a modern synthesis*, New York, St. Martin's Press, 1956; popular edition, New York, New American Library, 1957; see, for example, p. 24, p. 70-4. 'It is not possible for linguists to hand over phonetics to physics, and meaning to sociology, as some have proposed, without making structural linguistics utterly sterile, a risk of which this subject is already in great danger' (p. 135).
- 149 51. Erwin REIFLER, 'Linguistic analysis, meaning and comparative semantics', *Lingua*, 3, 1953, p. 371-90.
- 149 52. CECCATO began by a criticism of what he referred to as the 'cognitive tradition', see especially 'Il teocono', *Methodos*, I, 1949. He subsequently analysed mental operations, in particular in three important publications: *Il linguaggio, con la Tabella di Ceccatieff* [Language, with tables by Ceccatieff], Paris, Hermann, 1951; 'L'école opérationnelle et la rupture de la tradition cognitive' [The operational school and rupture of the cognitive tradition], *Bull. Soc. franc. philos.*, March 1952-May 1953; 'Contra Dingler, pro Dingler', *Methodos*, IV, 1952. This analysis immediately involved certain linguistic implications (see *Il linguaggio*, p. 26-8). The 'Ceccatieff table' (the 'Russification' of the Italian name of the author was intended evidently to indicate an analogy with Mendeleev's classification) provides a classification of 'traditional semantizations' corresponding to mental operations and to the results of such operations (p. 188-212 of *Il linguaggio*), according to four criteria. In 'Contra Dingler', Ceccato defines in operational terms, articles (p. 236), the verb (p. 237), the grammatical categories of subject and of intransitive mode, of object and of transitive mode (p. 262-3).
- 149 53. See report by S. CECCATO and Enrico MARETTI, 'Suggestions for mechanical translation', to the Symposium on the Theory of Information, London, 12-16 September, 1955, and published in Colin CHERRY (ed.), *Information theory*, p. 171-80, London, Butterworths Scientific Publications, 1956. This first publication was followed by five other articles, of which the most important is 'La traduzione meccanica' [Mechanical translation], *Automazione e automatismi* [Automation and automatism], II, 2, March-April 1958, p. 1-12; the report by S. CECCATO, E. MARETTI and E. ALBANI to the Cleveland Conference 'Classifications, rules, and code of an operational grammar for mechanical translation' is a detailed restatement of the methods followed, based on the idea of 'correlation' (note, in particular, the classification of 'correlators' given on p. 10-12). The reader may also read with profit the brochure *Adamo II, modello meccanico di operazioni mentali* [Adam II, mechanical model of mental operations], published upon the occasion of the International Congress on Automatism, Milan, 8-13 April 1956, and Ceccato's article, 'Tempo e spazio nella cibernetica' [Time and space in cybernetics], *Archivio di Filosofia*, 1958, where may be found perhaps the clearest outline of the general ideas which are the basis of the Ceccation thesis (p. 153-5); see also, especially, this

- definition of thought: 'To think is to correlate, and is to open and close correlations. The lengthier activity which is always represented by a mental category, becomes the correlating element, the activities of a lesser duration, which are either of a mental or other nature, become the correlated elements.'
- 149 54. CECCATO has shown to us a few of the preliminary materials relating to this. He participated actively in the recent Grisa Seminar in Brussels: see his contribution, 'I problemi filosofici del linguaggio', p. 37-50 of the collection *Enseignement préparatoire aux techniques de la documentation automatique*, Brussels, Euratom, 1961.
- 149 55. J. KURYŁOWICZ; 'Lexical and syntactic derivations', *Bull. Soc. Ling. Paris*, 37, 1936, p. 79-93; 'Fundamental structures of language: groups and propositions', *Studia philosophica*, 3, 1948, p. 203-9; 'The isomorphism notion', *Travaux Cercle ling. Copnh.* 5, 1949, p. 48-60; 'The problem of the classification of cases', *Biuletyn polskiego towarzystwa językoznawczego* [Bulletin of the Polish Society of Linguists], 9, 1949, p. 20-43. The principal articles by J. Kuryłowicz since the war have been brought together in a collection *Esquisses linguistiques*, Wrocław-Krakow, Zakład narodowy imienia Ossolińskich wydawnictwo polskiej Akademii Nauk, 1960; see Leon ZAWADOWSKI, *Constructions grammaticales et formes périphrastiques*, Krakow, 1959, and 'La signification des morphèmes polysèmes', *Biuletyn polskiego towarzystwa językoznawczego*, 17, 1958, p. 67-95.
- 149 56. J. KURYŁOWICZ, *Biul. pols. tow. Język.*, 14, 1955, p. 1-11. See also A. MIROWICZ' article on the concept of grammatical modality and the problem of particles, *ibid.*, 15, 1956, 81-92. We know of, by a J. Carroll reference only, Tadeusz MIŁEWSKI's work on the theory of linguistics (*Zarys językoznawstwa ogólnego, I: Teoria językoznawstwa* [Outline of general linguistics, I: Theory of linguistics], Lublin, Nakład i wydawnictwo, 1947).
- 149 57. J. HERMAN and I. PAPP, on 'thought and grammatical structure' and 'role of the grammatical function in the crystallization of the parts of speech', *Mélanges D. Pais (Emlékkönyv Pais Dezső)*, Budapest, 1956).
- 149 58. I. A. MEL'ČUK, *Mašinnyj perevod i prikladnaja lingvistika*, no. 2 (9), 1959, p. 59-69 (English translation by the US Joint Publications Research Service, JPRS 3599, p. 43-9).
- 149 58a. See Josef VACHEK, *Dictionnaire de linguistique de l'école de Prague*, Utrecht, Spectrum, 1960.
- 149 59. B. TRNKA, 'Prague structural linguistics', *Philologica Pragensia*, 1, 1958, p. 33-40. Professor Trnka showed us some interesting lectures given by him at the Charles University, unfortunately duplicated, and of which, it seems, there are no copies in the West.
- 149 60. Petr SGALL and Bohumil POLAK, *Slovo a slovenost*, 1959 (translated by the US Joint Publications Research Service, JPRS 3502, 13 July 1960).

- 149 61. L. SAINÉNU, *Reporturile intre grammatice si logica* [Relationships between grammar and logic], Bucharest, 1891.
- 150 62. See L. THOMAS, *The linguistic theories of N. J. Marr*, Berkeley, University of California Press, 1957.
- 150 63. V. Z. PANFILOV, 'Concerning the relationships between language and thought', published in 1957 in a handbook of the Institute of Philosophy of the USSR Academy of Sciences, and translated in *Recherches Internationales*, no. 7, May-June 1958, p. 74-93; see p. 84-6 and 92; the entire article, however, is very disappointing and misinformed.
- 150 64. 'For a discussion of problems of structuralism' (translated into French by the Société de linguistique de Paris, see *Bull. signalétique CNRS*, series on Philosophy, 1957, no. 2, p. 384-5).
- 150 65. It is interesting to note that the Association for Automatic Translation has set up a Seminar on Semiotics and Structural Linguistics, on which work began in March 1959. See the *Proceedings* (translated into English by the US Joint Publications Research Service, JPRS 3597, p. 2; JPRS 3758, p. 76-9; JPRS 8026, p. 57-8).
- 150 66. See I. I. REVZIN's report to the symposium held in Leningrad, 1-4 October 1957, on linguistic statistics: 'Relation between structuralist methods and statistical methods in modern linguistics', *Voprosy statistiki reči*, p. 45-57, Leningrad, L. R. Zinder, 1958 (p. 43-53 of the English translation, JPRS 6543).
- 150 67. I. A. MEL'ČUK gave at the symposium on linguistic statistics a remarkable report on 'Statistics of relation between terminations and gender of nouns in French' *Voprosy statistiki reči*, p. 112-30, op. cit. (p. 104-19 of the English translation, JPRS 6543). His very important paper at the Institute of Linguistics of the Academy of Sciences on 11 February 1958 on 'A model intermediary language for mechanical translation' was summarized in *Voprosy Jazykoznanija*, May-June 1958, p. 149 (translated in JPRS/DC-319, p. 25-6); a more detailed text reproducing his paper to the Conference on Mathematical Linguistics at Leningrad, 15-21 April 1959, appeared in *Mašinnyj Perevod i Prikladnaja Lingvistika*, no. 4, 1960, p. 25-45 (translated in JPRS 8026, p. 15-26).
- 150 68. See N. D. ANDREEV, 'The universal code of science and machine languages', paper to the conference at Cleveland (duplicated), and his article on 'Mechanical translation and the problem of an intermediary language', *Voprosy jazykoznanija*, no. 5, 1957, p. 117-21. Andreev was the originator of the Conference on Mathematical Linguistics held at Leningrad, of which—as far as we know—only abstracts are available in the West: *Tezisi soveščanija po matematičeskoj lingvistiki*, Leningrad, 1959. He signed, with Mel'čuk and V. V. Ivanov, a very interesting paper presented to the Cybernetics Section of the Conference on Mechanical Calculation on 17 November 1959, 'Some remarks and suggestions concerning work on mechanical transla-

- tion in the USSR', *Mašinnyj Perevod i Prikladnaja Lingvistika*, no. 4, 1960, p. 3-24 (translated in JPRS 8026, p. 1-14).
- 150 69. O. S. KULAGINA, 'On a method of definition of grammatical notions using the theory of sets', *Problemy kibernetiki*, no. 1, 1958, p. 203-14 (English translation, New York, Pergamon Press, 1959).
- 150 70. S. K. ŠAUMJAN, *Strukturnaja lingvistika kak immanentnaja teorija jazyka*, Moscow, Institute of Slav Studies of the Academy of Sciences, 1958. This pamphlet has been very fiercely attacked, notably by A. A. Leont'ev (see English translation in JPRS 6152, p. 7-14). Šaumjan in his turn severely attacked a very 'anti-structuralist' memorandum of the Committee for General Linguistics of the Academy of Sciences (published in June 1959) in a note which was published alone in February 1960 (English translation JPRS 6236, for the memorandum, see p. 1-11, and for the note by Šaumjan, p. 29-35). At the meeting on 7-8 July of the Division for Literature and Language of the Academy of Sciences, Šaumjan's theses were condemned (p. 40), and the decree of the Committee of this Division, passed at this meeting, denied to structuralism the right of entry into Marxist linguistics except indirectly through applied linguistics (ibid. p. 41-6, see especially p. 44-5). But in September 1960 *Voprosy filosofii* published an article by the same Šaumjan on 'Linguistic problems of cybernetics and structural linguistics' (English translation JPRS 6506, p. 120-31), which is a good restatement, without controversy.
- 150 71. It is interesting to observe that the 'modern' tendencies, bitterly criticized by traditional linguists, have nevertheless received support from acoustical experts—the Committee for Applied Linguistics was set up within the Acoustical Commission of the Academy of Sciences, see *Voprosy jazykoznanija*, no. 3, 1958, p. 136-7 (English translation JPRS/DC-319, p. 21-4)—and from experts in cybernetics—a linguistic section was set up on 3 July 1959 by the Scientific Council for the Co-ordination of Work on Cybernetics, see *Voprosy jazykoznanija*, no. 6, 1959, p. 150-1 (English translation JPRS 3597, p. 1). That the game may be considered to have been won by the 'moderns' would seem clear from the article by V. I. GRIGOR'EV, secretary of an *ad hoc* committee set up by the Academy of Sciences in February 1960 'The development of structural and mathematical methods in linguistic research', *Voprosy jazykoznanija*, no. 4, 1960, p. 153-5 (English translation JPRS 6732, p. 42-5), which comments on the decree of 6 May 1960 of the Praesidium of the Academy. At the first All-Union Conference on Mechanical Translation and Applied Linguistics of 15-21 May 1958 (see JPRS 1006-D, giving translations of the summaries which appeared in *Mašinnyj Perevod i Prikladnaja Lingvistika* 1 (8), 1959), structuralist methods were in fact quoted, and even approved, but it was not until the Conference on Applied Linguistics held at Černovcy, 22-28 September 1960 that a special section was devoted to

- structural linguistics and named after it, see *Voprosy jazykoznanija*, no. 1, 1961, p. 155-9 (translation JPRS 8132, p. 1-8).
- 151 72. Note particularly the passages (p. 112 and p. 114) indicating the opposition of the authors to a 'naturalistic attitude' towards language. See also the passage (p. 119-20) devoted to 'études grammaticales' which alludes to numerous polemics on the grammatical structure of the Chinese language. If one may judge from the published manuals—prepared, it is true, for the use of foreigners—by the University of Peking, the Chinese grammarians acknowledge the existence in their language of 'parts of speech' very similar to those of Occidental languages: nouns, pronouns, verbs, adjectives, numerals, prepositions, adverbs, conjunctions, interjections, with the necessary addition of 'measurers' and of 'particles' (see the *Jy fa giau cai* [Textbook: Rules of grammar], 1953, translated into German by Martin Piasek, Leipzig, Harrassowitz, 1957; and the *Modern Chinese reader*, II, p. 695-9, Peking, Epoch, 1958). Nothing is farther from Brøndal's conception (*Les Parties du discours*, p. 169-71, op. cit.). As to the Chinese 'monosyllabism', it is justly rejected, and numerous dissyllables are found, as well as words of 3 or 4 syllables (*Modern Chinese reader*, I, 21-2).
- 151 73. Hsu KO-CHANG and I. M. OSANIN, 'A survey of structural linguistics', originally published in *Si-fan Yu-yen*, no. 2, 1958 (English translation JPRS 992-D-II from the Russian translation by SEROV, *Voprosy jazykoznanija*, no. 3, 1959, p. 41-60).
- 151 74. *Voprosy jazykoznanija*, no. 5, 1959, p. 102-4, published a Chinese paper given at the Leningrad Conference on Mathematical Linguistics of 5-21 April 1959, which was translated into English (JPRS 1131-D); one of the articles quoted in this paper (that by Liu Yung-ch'uan on the problem of the order of words for mechanical translation from Russian to Chinese, published in *Yu-yen yen-chiu*, no. 4, 1959) has been translated into English (JPRS 3356).
- 151 75. Seventh International Congress of Linguists, *Proceedings*, report, p. 29-34; contributions, p. 35-45; discussion, p. 251-97.
- 151 76. *Ibid.*, BUYSENS, p. 35; HAAS, p. 40; HAUDRICOURT, p. 40; MATTHEWS, p. 43.
- 151 77. *Ibid.*, FUNKE, p. 251-73; see more particularly 'Modern attempts at classification' (p. 258-71).
- 151 78. *Ibid.*, report by COLLINSON, p. 65; contributions of ELLIS and HALLIDAY, p. 71, HAMMERICH, p. 73-4; discussion, p. 333-9; the most interesting intervention was that of Halliday, p. 336-8.
- 151 79. Gerlach ROYEN, *Die nominalen Klassifikations-Systeme in den Sprachen der Erde* [Nominal classification systems in the languages of the world], Mödling, Vienna, Anthropos, 1929.
- 151 80. See M. Cohen, *Pour une sociologie du langage*, p. 152-7, op. cit., with rather numerous references up to 1954. To these can be added: G. L. HALL and I. S. CLAIR-SOBELL, 'Animate gender

in Slavonic and Romance languages', *Lingua*, 4, 1954, p. 194-206; James E. IANUCCI, *Lexical number in Spanish nouns, with reference to their English equivalents*, Philadelphia, University of Pennsylvania, 1952; Robert A. FOWKES, 'Gender redistribution in Keltic', in *Studies . . . Whatmough*, p. 39-46, 1957; A. MARTINET, 'Le genre féminin en indo-européen', *Bull. Soc. Ling. Paris*, 1956, p. 83-95 (in contrast to the older Meillet thesis, interprets the formal opposition of gender as being born of 'necessities of agreements').

- 151 81. Apart from G. Guillaume's works (Note 24) and of R. Jakobson (Note 49); see, among others, E. LOCKER, 'Être et avoir, leur expression dans les langues', *Anthropos*, 49, 1954, p. 481-520; Martin Sánchez RUIPÉREZ, *Estructura del sistema de aspectos y tempos del verbo griego antiguo; análisis funcional sincrónico* [Structure of the system of aspects and tenses in ancient Greek: functional synchronous analysis], Salamanca, CSIC, 1954; J. M. BUFFIN, *Remarques sur les moyens d'expression de la durée et du temps en français*, Presses Universitaires de France, 1925); R. L. WAGNER, 'Verbes, préfixes, adverbes complémentaires', *Études . . . Mario Roques*, p. 207-16, 1946; Hans WEBER, *Das Tempussystem des Deutschen und des Französischen* [The tense system in German and French], Bern, Francke, 1954; Paul IMBS, *Le subjonctif*, Strasbourg, 1953; William E. BULL and Rodger FORLEY, 'An exploratory study of the nature of actions and the functions of verbs in Spanish', *Hispania*, 32, 1949, p. 64-73; H. MARCHAND, 'On a question of aspect: a comparison between the progressive form of English and that in Italian and Spanish', *Studia linguistica*, 9, 1955, p. 45-52; H. J. J. M. VAN DER MERWE, *et al.*, 'Aspek as uitdrukkingmiddel van handelings', *Mededelings van die Univ. van Suid-Afrika*, Pretoria, 1958; 'L'aspect verbal', by various authors, *Revue belge de philologie et d'histoire*, 1958, p. 118-43 and p. 871-6. This list, of course, could be extended indefinitely, and consists of merely a few studies which appeared to us, for one reason or another, to be of particular interest. Marcel COHEN's thesis *Le système verbal sémitique et l'expression du temps*, Paris, Leroux, 1924, is a work which was epoch-making and of which the general scope extends beyond the framework of Semitic languages alone; two studies by the same author, in rather significant detail will be found in *Cinquante années de recherches*, p. 194-205 and p. 227-47, Paris, Klincksieck, 1955. In BRØNDAL'S *Essais . . .*, there is a study of 1942 relating to 'The fundamental forms of the verb' (p. 128-33) in which he applies his general theory of definition by combination of generic terms. See, Note 35, the reference to Hjelmslev's research work on the verb and the noun phrase.

- 151 82. Paul FORCHHEIMER, *The category of person in language*; we have not seen this study, which has come to our attention solely as a result of S. NEWMAN'S review in *American Anthropologist*, 56, 1954, p. 926-7. See also: V. BRØNDAL'S 'Le concept de

- "personne" en grammaire', *J. Psychol.*, 1939, p. 175-82 reproduced in the *Essais . . .*, 1943, p. 98-104; L. HJELMSLEV, 'La nature du pronom', *Mélanges . . . Van Ginneken*, 1937, p. 51-8; Pier ERINGA, 'Les pronoms des langues classiques et la morphologie moderne', *Lingua*, 3, 1952, pp. 69-97.
- 151 83. See above, Notes 36 and 49. See also: Hans VOGT, 'L'étude des systèmes de cas', *Travaux Cercle ling. Copnh.*, 5, 1949, p. 112-122; Hand Christian SORENSEN, 'Contribution à la discussion sur la théorie des cas', *ibid.*, p. 123-33; Thomas A. SEBEOK, *Finnish and Hungarian case systems*, Stockholm, 1946. The suffixing system of Hungarian names to express different relationships (aggregation, dependency, participation, 'address') has been studied by Janos LOTZ in *Travaux Cercle ling. Copnh.*, 5, 1949, p. 185-197, with the aid of an interesting graphic diagram.
- 151 84. Question A3. COLLINSON report, p. 63-4; contributions (ERINGA, ELLIS, GUTHRIE, POTTIER), p. 66-70; discussion, p. 315-29. See the already older works of O. JESPERSEN, 'Negation in English and other languages', *Kgl. Danske Vidensk. Selskab, Hist.-filol. Medd.* [Proceedings of the Royal Danish Academy of Sciences: history of philology] vol. I, p. 5, Copenhagen, 1917, and of A. ZOBEL, *Die Verneinung im Schlesischen* [Negation in the Silesian dialect], Breslau, 1928.
- 151 85. CARROLL, *op. cit.*, p. 127 (see Note 38 above).
- 152 86. E. SAPIR, *Totality*, Linguistic Society of America, 1930 (Language monographs, no. 6) (the introduction provides a plan of the 15 sections which were to comprise the complete series, under the title 'Foundations of language, logical and psychological: an approach to the international language problem'); 'Grading, a study in semantics', *Philosophy of science*, 11, 1944, 93-116, reproduced in *Selected writings of Edward Sapir*, 1949, p. 122-149. E. SAPIR and MORRIS SWADESH, *The expression of the ending-point relation in English, French, and German*, 1932 (Language monographs, no. 10). William Edward COLLINSON, *Indication, a study of demonstratives, articles, and other 'indicators'*, 1937 (Language monographs, no. 17). Concerning totality, see also V. BRØNDAL's contribution 'Omnis et totus, analyse et étymologie', *Mélanges . . . Pedersen*, 1937, reproduced in the *Essais . . .*, p. 25-32, and Karl BRUGMANN, *Die Ausdrücke für den Begriff der Totalität* [Expressions for the concept of totality], (programme of the University of Leipzig, 1894). K. SNEYDERS DE VOGEL's study on *Les mots d'identité et d'égalité dans les langues romaines*, Wageningen, 1947.
- 152 87. V. BRØNDAL, 'Théorie de la dérivation', written in 1942, published in the *Essais . . .*, p. 124-7; C. AYMONIER, *Essai sur la dérivation comparée dans les langues naturelles et artificielles*, Paris, 1921. On the French language: Arsène DARMESTETER, *Traité de la formation des mots composés dans la langue française*, Paris, Bouillon, 1894, 2nd ed. revised by Gaston, Paris (the 1st ed. is dated 1875); by the same author *De la création actuelle de mots*

nouveaux dans la langue française, Paris, 1877; E. PICHON, 'Les principes de la suffixation en français', articles in *Le Français moderne*, 1937-40; A. DAUZAT, 'L'appauvrissement de la dérivation en français', *ibid.*, 5, 1937, p. 289-300; J. LÉGER, 'A propos du préfixe re-', *ibid.*, 24, 1956, p. 285-91. See the very remarkable study by J. GREENBERG on the order of suffixes, already cited in Note 45 above. HANS MARCHAND published an important book on *The categories and types of present-day English word-formations*, Wiesbaden, Harrassowitz, 1960.

- 152 88. See in the *Proceedings* of the Seventh International Congress of Linguists, in the first plenary session, 'Linguistics and the problem of meaning'; report by FIRTH, p. 5-9; contributions, p. 10-17 and p. 178; discussion, p. 181-233. At the same congress, session IIb, note the observations made by O. Funke concerning the necessity of considering semantic aspects, p. 263-5. As Marcel Cohen reminded the seventh congress, p. 185-6, one should of course consider also the semantic aspect in phonology, but this problem is of no direct concern to us here.
- 152 89. One would be mistaken in believing that this is a recent trend, since, as Franz DORNSEIFF reminds us in his remarkable 'Vorrede' [Preamble], *Der Deutsche Wortschatz nach Sachgruppen* [The German vocabulary by subject groups], Berlin, W. de Gruyter, 1934, one finds tentative systematic vocabularies at Babylon in the third millennium before Christ. In modern Europe, the most important work on systematic lexicology was that of Peter Mark ROGET, in the nineteenth century, the *Thesaurus of English words and phrases*, which he was in process of preparing as early as 1806, and of which the first edition appeared in 1852; see also among numerous contemporary editions that of Penguin Books, London, 1953. Concerning Roget, see Henry SWEET, 'Words, logic, and grammar', *Trans. Philological Soc.*, 1875-76, p. 470-503, reproduced in his *Collected papers*, p. 1-33. Oxford, Clarendon Press, 1913. The seventh International Congress of Linguists had put 'conceptual dictionaries' on its agenda (point A5): see F. MEZGER's report, p. 77-85, the contributions p. 86-9, and the discussion p. 443-73. One of the most remarkable ideological dictionaries was that of J. CASARES, *Diccionario ideológico de la lengua española* [Ideological dictionary of the Spanish language], Barcelona, 1942; see also his *Introducción a la lexicografía moderna*, Madrid, 1950. Under the impulsion of Antoine Thomas, a certain number of French dialectological studies were made on a systematic basis, e.g. L. LHERMET, *Contribution à la lexicologie du dialecte aurillacois*, Paris, 1931. Walther VON WARTBURG constituted himself the protagonist of the general application of such a method, see his report 'Das Ineinandergreifen von deskriptiver und historischer Sprachwissenschaft (1931); 'Betrachtungen über die Gliederung des Wortschatzes und die Gestaltung des Wörterbuchs' [The interaction of descriptive and historical linguistics (1931); views on the structure of the vocabulary and the formation of the dictionary],

Mélanges Bally, 1939; *Problèmes et méthodes de la linguistique*, p. 159-62 Paris, Presses Universitaires de France, 1946; Rudolf HALLIG and W. VON WARTBURG, 'Begriffssystem als Grundlage für die Lexikographie; Versuch eines Ordnungsschemas' [The system of concepts as the foundation for lexicography; a tentative system of arrangement], *Abhandlungen der deutschen Akademie der Wissenschaft zu Berlin, Klasse für Sprache, Literatur und Kunst* [Proceedings of the German Academy of Sciences in Berlin, Class for language, literature, and art], 1952, no. 4. See on this subject W. RUNKEWITZ, 'Kritische Betrachtungen zum Begriffssystem von Hallig v. Wartburg im Zusammenhang mit den Arbeiten am Altgaskognischen Wörterbuch' [Critical remarks on the system of concepts of H. v. Wartburg in connexion with the work on the Old Gascon dictionary], *Monatsberichte der Deutschen Akademie der Wissenschaften zu Berlin*, 1946-56, p. 379-88, which we have not seen. See also recently F. DE TOLLENAERE, 'Lexicographie alphabétique ou idéologique', *Cahiers de lexicologie*, no. 2, 1960, p. 19-29. We have already mentioned (Note 11 above) the German research on 'semantic fields', which later inspired Georges MATORÉ, *La méthode en lexicologie, domaine français* [Method in lexicology in the field of the French language], Paris, Didier, 1953; he offers here (p. 70-4) a diagram of 'a comprehensive classification of lexicon facts' different from that of Hallig and Wartburg, and moreover, less satisfactory. It will be noted that Matoré, in defining 'lexicology as a sociological discipline using words as its linguistic material', tries to make of it an 'autonomous discipline', the field of which partly covers that of linguistics, but independently of it: for reasons other than those of certain American structuralists, this position results in a dismemberment of linguistics in a way which does not seem to be any longer justifiable (p. 50-1). It does not seem essential here to mention American 'general semantics', developed around the writings of Korzybski, and later of Hayakawa; see the critical passage devoted to it by CARROLL (op. cit., p. 164-8, see Note 38 above). Concerning structural semantics in general, and its (desirable) relationships with other parts of structural linguistics, see S. ULLMANN in the second edition of his *Principles of semantics*, p. 307-21, op. cit., with numerous references, and Uriel WEINREICH, 'On semantic universals' (duplicated, 71 pp., March 1961, with an important bibliography), and also his programme of studies, 'Semantic structure of natural languages' (duplicated memorandum, 5 May 1961). At the eighth International Congress of Linguists (Oslo, 1957) there was a (rather disappointing) discussion on the subject 'To what extent can meaning said to be structured?' (p. 636-704 of the *Proceedings*), of which the most interesting item was the paper by Hjelmslev, which we have already mentioned. See also Hans POLLAK, 'Gibt es Wortklassen vom Standpunkt der Bedeutung?' (Are there word-classes from the point of view of meaning?), *Beiträge zur Geschichte der deutschen Sprache und Literatur* (Tübingen), 80, 1958, p. 33-47. We have not seen:

- Andrew Paul USHENKO, *The field theory of meaning*, Ann Arbor, University of Michigan Press, 1958; F. F. NESBIT, *Language, meaning and reality*, New York, 1955; H. R. WALPOLE, *Semantics*, Norton, 1941. On basic concepts, see Eric H. LENNEBERG and John M. ROBERTS, *The language of experience*, supplement to the *International journal of American linguistics*, 22, (2) Charles E. Osgood and his collaborators have tried to apply a method called the method of 'semantic differential', to obtain a 'measurement' of meaning: see C. E. OSGOOD, George J. SUCI and Percy H. TANNENBAUM, *The measurement of meaning*, Urbana, University of Illinois Press, 1957; but Uriel WEINREICH has rightly observed in 'Travels through semantic space', *Word*, 14 (2-3), 1958, p. 346-66) that the 'semantic differential' measures 'meaning' only in a psychological sense, from the point of view of the emotional reactions of the subjects studied to such and such a word (cf. especially p. 358-60 of his article).
- 152 90. The question was: 'Can a logical calculus be devised that shall be structurally independent of the grammatical patterns of any language?' See *Proceedings*: E. BUYSSENS' report, p. 21-2; contributions (HAAS, WHATMOUGH), p. 23-26; discussion, p. 237-48 (in which Haas, p. 242, observes that the 'logical syntax of language' of Carnap is in reality a syntax of logical language, and his 'Introduction to semantics' a semantics of logical language). This problem called to mind, in fact, the debates for or against logical positivism of the Vienna Group, the Polish school of Tarski, Wittgenstein, etc., and on which a bibliography will be found in S. ULLMANN, *The principles of semantics*, op. cit., p. 302-4 (see also his observations on logic and language in general, p. 12-16), the size of which could easily be tripled (see, among others, Gustav BERGMANN's recent handbook, *Meaning and existence*, Madison, University of Wisconsin Press, 1960; Albert SHALOM's article, 'Wittgenstein, language and philosophy', *Études philosophiques*, 13, p. 486-96).
- 152 91. The best is perhaps that of Marcel COHEN in his article 'Faits linguistiques et faits de pensée', *J. psychol.*, 1947, p. 385-402, reproduced in *Cinquantes années de recherches*, 25-37 (see p. 31-36); however, many of the formulae would still need to be discussed. See also Ch. SERRUS, *Le parallélisme logicogrammatical*, Paris, Presses Universitaires de France, 1933, and, by the same author, *La langue, le sens, la pensée*, Paris, Presses Universitaires de France, 1941; these two works are discussed in G. ANTOINE's thesis, *La coordination en français* p. 87-111, (mentioned earlier in Note 20).
- 152 92. For example, Emile BENVENISTE's very remarkable article, 'Catégories de pensée et catégories de langue', *Études philosophiques*, 13, 1958, p. 419-29, where he proves that 'Aristotle's list of ten categories can be transcribed in terms of language'. See also: the brief paper by E. W. COUNT 'Symposium: do we need more becoming words', *American anthropologist*, 55, 1953, p. 395-403; two articles by Y. R. CHAO, 'The logical structure of Chinese

- words', *Language*, 22, 1946, p. 4-13, and 'Notes on Chinese grammar and logic', *Philosophy East & West*, 5, 1955, p. 31-41 (with a complementary note by H. H. DUBS, p. 167-8). Concerning the more general problem of language and philosophy: a few observations by LÉON BRUNSCHVICG, *Héritage de mots, héritages d'idées*, p. 16, 21 and 69, Paris, Presses Universitaires de France, 1945; Antony FLEW 'Philosophy and language', *Philosophical quarterly*, 1955 (reproduced with correspondence as chapter I of the collection *Essays in conceptual analysis*, 1956, p. 1-20); the collection of *Essays on logic and language*, London and New York, 1951 (published under Flew's direction); the interesting (but rather disappointing) investigation by A. H. BASSON and D. J. O'CONNOR, 'Language and philosophy', *Philosophy*, 22, 1947, p. 49-65; P. BELOV, 'Sur la langue dans le processus de la connaissance', *Actes of the 2nd International Congress of the International Union of Philosophy of the Sciences, Zürich, 1954*, vol. III, p. 131-6, Neuchâtel, Griffon editions, 1955.
- 152 93. In the 1956 collection mentioned above in Note 44, this 'hypothesis' is seen to appear as early as 1936 (article on the Hopi verbs, p. 51-6, and, especially, the article 'An American Indian model of the universe' also of 1936, but published only in 1950; see p. 57-64; also in a 1936 text, unpublished until then, p. 83-4). It probably finds its best expression in the report 'The relation of habitual thought and behavior in language' written in 1939 and published in 1941 in the volume in memory of Sapir (p. 134-59 of the 1956 collection; see more particularly p. 158); Whorf popularized it, finally, in two articles in the *Technology Review* of the Massachusetts Institute of Technology for 1940-41 (p. 207-19—see among others, on p. 214-5, the formula of a 'new principle of relativity: the relativity of all conceptual systems, ours included, and their dependence upon language'—and p. 233-45) and in a sort of spiritual testament published the year of his death (1942) in the *Theosophist* (p. 246-70; especially the conclusion, p. 269-70). The reading of Whorf's book is extremely stimulating, anyway.
- 152 94. In the first instance Harry HOMER: see his article 'Cultural implications of some Navaho linguistic categories', *Language*, 27, 1951, p. 111-20; 'The Spari-Whorf hypothesis', p. 92-105 of the collection *Language in culture*, University of Chicago Press, 1954, published under his direction, containing papers given at a conference convened in 1953 for the particular purpose of discussing the Whorfian theses. L. von BERTALANFFY, 'An essay on the relativity of categories', *Philosophy of science*, 22, 1955, p. 243-63.
- 152 95. For example, J. WHATMOUGH, *Language: a modern synthesis*, op. cit., p. 84 and p. 200-3 (see Note 50). It will be observed that M. COHEN, in his 1947 article (cited above in Note 91) takes almost an exactly opposite view of Whorf's thesis (whom he probably did not know at the time). Many linguists participating in the 1953 conference maintained a reserved attitude: for

instance HOCKETT (*Proceedings* p. 122-3 and p. 128) and GREENBERG (p. 130-1. Cf. LÉVI-STRAUSS (op. cit., p. 84) writes that 'Whorf attempts to discover correlations between objects derived from two very distant levels, by the quality of observation and the refinement of analysis to which one and the other are subjected' and suggests a comparison between the Hopi linguistic structures studied by Whorf and those of the systems of kinship in the same society (p. 84-7).

- 153 96. CARROLL's conclusion (*The study of a language . . .*, op. cit., p. 46) was that Whorf was by no means the first to outline such ideas; as noted by Carroll (ibid., p. 45) some of them can be found in Humboldt and also in Jan BAUDOIN DE COURTENAY (*Einfluss der Sprache auf Weltanschauung und Stimmung* [Influence of language on philosophy of life and frame of mind], Warsaw, 1929); see also the statement by OGDEN and RICHARDS, *The meaning of meaning*, London, 1923, p. 195: 'so far from a grammar—the structure of a symbol system—being a reflection of the world, any supposed structure of the world is more probably a reflection of the grammar used', to which BRONDAL adhered (*Parties du discours*, p. 35). See also the article by J. FOURQUET 'Pensée et grammaire', *Études philosophiques*, 13, 1958, p. 430-45, of which certain formulae are moreover very disputable, such as that (p. 440, note) where he writes that 'the category of the *general*, which seems inseparable from the notion of noun, seems to be lacking in a language such as the Chinese', or (p. 444—it is true, however, leaning on H. Maspero) that the 'Chinese or Vietnamese . . . encounter more difficulty in *generalizing* than the Europeans, which could be due to the nature of the isolating linguistic type'. One can only maintain an attitude of extreme mistrust in the face of such affirmations; it would certainly be wise to meditate in this connexion on Meillet's observation on the distinction between a category and such or such a form by which it is expressed (conclusion of the *Esquisse de la langue latine*, 1928, cited in M. COHEN's *Pour une sociologie du langage*, op. cit., p. 158). Hans HARTMANN attempted to relate verbal forms and the religious concepts to which they would correspond: *Das Passiv: eine Studie zur Geistesgeschichte der Kelten, Italiker und Arier* [The passive voice; study of the intellectual history of the Celts, Italians and Aryans], Heidelberg, Winter, 1954; see also his observations at the seventh Congress of Linguists, *Proceedings*; p. 509-13.

- 153 97. M. COHEN, *Pour une sociologie . . .*, op. cit., p. 17-24; J. B. CARROLL, *The study of a language . . .*, op. cit., p. 69-111 (one of the best in his book); J. VAN GINNIKEN, *Principes de linguistique psychologique*, Paris: Rivière, 1907; J. LAROCLETTE, 'La psychologie des peuples et l'étude du langage', *Revue de psychologie des peuples*, p. 188-209, 1952; Friedrich KAINZ, *Psychologie der Sprache* [The psychology of language], Stuttgart-Vienna, 1941-56, 4 vols.; S. ULLMANN, *The principles of semantics*, op. cit. (see the index under 'Psychology'); Franklin FEARING, 'An examination of the conceptions of Benjamin Whorf in the light of theories of

- perception and cognition', In: H. HOJER, *Language in culture*, op. cit., p. 47-81—and the references cited by these authors.
- 153 98. André OMBREDANE, *L'aphasie et l'élaboration de la pensée explicite* (Paris, Presses Universitaires de France, 1951)—the best synthesis to that date; see in particular what he writes concerning the 'categorical attitude', p. 251-4, p. 274-5, 370-1 and p. 380.
- 153 99. M. COHEN, *Pour une sociologie...*, p. 146-50; op. cit., B. MALINOWSKI, 'The problem of meaning in primitive languages', In: OGDEN and RICHARDS, *The meaning of meaning*, op. cit. (Supplement I), reproduced in *Magic, science, and religion*, p. 228-76, Boston, Beacon Press, 1923 (see CARROLL's critical observations, op. cit., p. 116 and p. 239-40). Concerning L. LÉVY-BRUHL, see *Revue Philosophique*, no. 4, vol. 82, 1957. In a recent article on the 'lexicology of so-called primitive peoples', K. MOSZYNSKY again draws attention to the abundance of names of concrete objects found therein (*Biul. Polsk. tow. Jezyk.*, 15, 1956, p. 93-112). See also: A. SPIRKIN, 'La formation de la pensée abstraite aux premiers stades du développement humains', translated from *Voprosy Filosofii*, 1954, in *Recherches soviétiques*, section I, 1956, p. 59-81; V. V. BUNAK, 'L'origine du langage', *Colloques internationaux du CNRS, Les processus de l'homínisation*, p. 99-111, Paris, CNRS, 1958. See also William THALBITZER's interesting communication, 'Is Eskimo a primitive language?', *Actes of the Fourth International Congress of Linguists*, 1936, p. 254-62.
- 153 100. In the introduction to the first edition of *Les langues du Monde*, 1924, MEILLET wrote: 'The only linguistic classification of any value and utility is the genealogical classification, founded on the history of languages' (*LHLG*, II, p. 53).
- 153 101. See the bibliography up to 1951 in the second edition of *Les langues du Monde* by MEILLET and COHEN, p. XXXIII-XXXV. The best comprehensive study (to our knowledge) is that of Emile BENVENISTE in his lecture at the Institute of Linguistics of the University of Paris, in January 1952, 'La classification des langues', published in the *Conférences* of the institute (vol. XI, 1952-53, p. 33-50); one should observe specially the indications given by him on the conditions of method to be observed in order to arrive at a more scientific classification by types, p. 47 and p. 49-50. Other recent references: J. LOHMANN, 'Sprachgeographie und Sprachtypologie' [Geography of languages and typology of languages], *Lexis*, 4, 1954, p. 87-98; C. E. VORGELIN, 'On developing typologies and revising old ones', *Southwest Journal of Anthropology*, 11, 1955, p. 355-60; L. TESNIÈRE, 'La classification par le sens du relevé linéaire', *Éléments de syntaxe structurale*, 1959, p. 32-3; Joseph H. GREENBERG, *Essays...*, p. 66-7, p. 73 and chapter VIII to which we referred earlier (Note 45)—it will be observed that he emphasizes the 'need for some cataloguing of facts regarding all languages that would permit a reliable answer based on systematically assembled data' (p. 89) and the fact that 'much remains to be done both in the accumulation of descriptive and

historical knowledge and in the codification and archiving of existing linguistic information' (p. 93); C. E. BAZELL, *Linguistic typology*, London, School of Oriental and African Studies, 1958; P. G. GANESH-SUNDARAM, 'A generalized treatment of the typology of languages as communicable media, *Bull. Deccan College Research Institute, Poona*, 12, (1951-52) p. 415-36. We have not seen P. S. KUNZNECOV's book, *Die morphologische Klassifikation der Sprache* [The morphological classification of language], nor even DEL NEGRO's review in the *Philos. Lit. Anzeiger*, 10, no. 1, 1957, p. 11-12, of which we have merely a reference. Certain interesting indications will be found in the *Proceedings* of the seventh Congress of Linguists, p. 55-9 and p. 301-12 (principles to be followed in the formal analysis of invariable word languages), and p. 121-30 and p. 439-59 (areas of grammatical affinity). Concerning W. Schmidt's attempt at establishing linguistic areas by types in *Die Sprachfamilien und Sprachenkreise der Erde* [The language families and linguistic regions of the earth], Heidelberg, 1926, see M. COHEN's review, *Bull. Soc. Ling. Paris*, 28, 1928 p. 10-21, of which extracts may be found in *Pour une sociologie . . .*, op. cit., p. 150-2. Ideas of Cl. LÉVI-STRAUSS, op. cit., p. 71-4 and p. 87-90, concerning the possibility of comparing areas of linguistic structure with areas of relationship systems. To be noted also is an idea of W. VON WARTBURG's, who suggests (*Problèmes et méthodes . . .*, op. cit., p. 120) 'comparing different languages from the standpoint of the distribution of words among three classes', which he distinguishes from the point of view of their more or less great 'motivation' (in that connexion he also gives an interesting indication concerning the problem of the 'arbitrariness of the symbol', which we have not discussed here, and on which see, among others, Cl. LÉVI-STRAUSS, op. cit., p. 103-8).

- 153 102. Paul MENZERATH, 'Typology of languages', *Journal of the Acoustical Society of America*, 22, 1950, p. 698-701; *Die Architektonik des deutschen Wortschatzes* [The structure of the German vocabulary], Bonn, Dümmler, 1954.
- 153 103. A suggestion in this connexion in Cl. LÉVI-STRAUSS, op. cit., p. 65-6. *Mutatis mutandis*, the method advocated by J. C. Gardin for the topological study of the forms of vases, note no. 1 of *Trois projets d'analyse structurale* (see Notes 29 and 37 of Chapter 2 of the present study) could be applied to this problem. We will not dwell here on the numerous preliminary studies of methodology that such an undertaking would necessitate.
- 153 104. Comprehensive outline of the question in O. JESPERSEN, 'Progress or decay?', *Language*, p. 319-36, London, Allen, 1922. Jespersen, himself, answers the question in the following chapter, entitled 'Progress' (p. 337-66)—but he considers here (see p. 364) only the 'ancient and modern languages belonging to our family of speech'. The three following chapters are also interesting (p. 367-442) and the conclusion is clearly optimistic: 'from the beginning the tendency has been one of progress, slow and fitful

progress, but still progress towards greater and greater clearness, regularity, ease, and pliancy' (p. 441-2).

154

105. In his opening Lesson, in 1906, MEILLET desires to determine 'not historical laws any longer . . . but general laws . . . for all times; which . . . will extend equally to all languages' (*LHLG*, I, p. 11); he mentions 'the progressive elimination of complex inflexions' and that 'of too complex forms' in the Indo-European languages (p. 12) and adds that this family 'certainly provides a vast enough field of observation for the conclusions to lay claim to a general validity' (p. 13); there remains only to seek, in all the other families of languages, the 'different conditions' and the 'variety of general conclusions offered for study', and which 'permit a verification of the value of general conclusions which can only be drawn from a study of the Indo-European languages' (p. 14). He practically never varied subsequently in his fundamental thought on this point: the same basic ideas are encountered again and again, like a *leitmotiv*, throughout the entire length of the two volumes of *Linguistique historique et linguistique générale*. See for example: vol. I, p. 40-1 (simplification of inflexion, 'reductions, thanks to which recent morphologies have attained the degree of simplicity and regularity' of Gothic and Armenian) with a somewhat analogous development in vol. II, p. 116-9 (where one notices, however, an idea previously only slightly developed, concerning the 'natural' causes of the persistence of verbal inflexion); vol. I, p. 66-71 ('everywhere the progress of civilization tends to destroy semi-concrete forms', p. 66; 'an effort to proceed from the form-word . . . to the existing word in isolation and resembling no other word', p. 68; however, he recognizes, concerning verbs, that 'as fast as the inflexion is destroyed so it tends to re-form itself', p. 71); vol. I, p. 155-8 (concerning the elimination of the simple preterite 'a moment of great development which encourages the Indo-European languages to proceed from the variable form-word to a word fixed once and for all'); vol. I, p. 193-8 (elimination, 'in the process of which, with progress of civilization, the categories take on a more abstract character', the 'categories having a concrete or expressive value').

In his great article of 1919 'the grammatical gender and elimination of inflexion', it is clear that Meillet, in order to remain faithful to his theory, stretched the fact slightly. For example: when he discusses the 'constant difference of gender peculiar to the pronoun' in English, which in his opinion constitutes 'a definite progress' (vol. I, p. 206)—neglecting the complexities and subtleties of the English gender, to which Whorf (p. 68-9 and 90-2 of the 1956 handbook), after Sapir (p. 159 of the French translation of *Language*) so rightly draws attention; or he interprets the facts rather hastily (the 'archaism of the Slavic languages' and 'the archaizing tendency of its development' being considered as 'consequences of the backwardness of the Slavs in relation to universal civilization', vol. I, 207); noting, however, that the Poles developed 'distinctions of gender' at a

time when the 'influence of Western civilization was in full sway', he explains the latter by a formula, which no doubt Whorf would have appreciated, 'Languages serve to express the mentality of the people speaking them, but each language constitutes a strongly organized system which imposes itself upon them, gives their thoughts its form and which only undergoes the influence of this mentality in a slow and partial manner, and as occasions present themselves'.

The outline presented in 1922 to the Society of Psychology, on 'The concrete character of the word' (vol. II, p. 12), revives the ideas contained in the note offered to M. Bréal in 1900 (vol. I, p. 40) and in the 1918 article (vol. I, p. 68); but a few students (Delacroix, Vendryes, Lalande) discuss Meillet's thesis rather vehemently (p. 13-21).

To give him his due, however, it must be noted that in Meillet there appear two other notions: that of a 'spiral development' (vol. I, p. 140-1) induced by the need for expressiveness (also vol. I, p. 163-7); and that of 'survival' (vol. I, p. 228; vol. II, p. 88-9) which is found in its clearest form in 'La morphologie est le domaine de la survivance', *La méthode comparative en linguistique historique*, p. 91-2, Oslo, Aschehoug, 1925.

- 154 106. Edward SAPIR, *Le langage*, p. 155 (French translation).
- 154 107. *Ibid.*, p. 157.
- 154 108. *Ibid.*, p. 160.
- 154 109. *Ibid.*, p. 205. Similar ideas, p. 28-9.
- 154 110. *Ibid.*, p. 131; in the same sense, critical remarks on 'form for the sake of form' (p. 96) (with a development which strongly recalls the ideas of Meillet on 'survivals'), and also 'the form which survives its conceptual content' (p. 94) 'illogical complications of our European languages . . . a system which so completely confuses the subject matter of speech with its form' (p. 93). The most perfect prototype of languages of 'conceptual type A'—simple languages with pure syntactic relationships—is the Chinese language. Modern English tends to come closer—by the combined action of the three evolutionary trends described by Sapir—to this type (which it is still far from reaching). This fact may have played a part in the genesis of Sapir's ideas on the question. It will be noted that, in the index to Bloomfield's *Language*, neither the word 'evolution' nor—a *fortiori*—the word 'progress' appear; only the neutral word 'change' is found.
- 154 111. Charles F. HOCKETT, *A course in modern Linguistics*, op. cit. 'Objective measurement is difficult, but impressionistically [!] it would seem that the total grammatical complexity of any language . . . is about the same as that of any other. This is not surprising, since all languages have about equally complex jobs to do [?], and what is not done morphologically has to be done syntactically' (p. 180-1). He freely admits that languages can be compared according to their 'average degree of morphological complexity' but adds immediately (without, however, a shadow

of evidence in support) 'carrying with it an inverse implication as to degree of syntactical complexity'. After having adopted as the 'easiest rough measure of morphological complexity' the average number of morphemes per word in a representative sampling of the language (which is, indeed, a very 'rough' measure!) he declares: 'There is no discernible correlation between the placement of a language on the analytic-synthetic scale and anything else about either the language or other aspect of the life of its speakers' and then rises against the 'false theory' which was 'that in course of time all languages tend to become increasingly analytic' (giving, moreover, as a sole argument that the French language is 'somewhat more synthetic now than it was a few centuries ago'—a thesis which could at least be debated). All of the foregoing sequence of statements brings him to his conclusion: 'A further misconception, a sort of corollary of that just mentioned, was the notion that some languages of today . . . are more "progressive" than others . . . because they have developed further in the analytic direction. For this there is no evidence at all'—which he 'proves' simply by mentioning that the Turks and the Chinese 'manage all the business of everyday life, and the complexities of modern technology' despite the differences in their languages from the standpoint of 'analytic-synthetic' character (p. 181-2). This amounts to disposing of the problem by a flick of the hand. Elsewhere, Hockett ('Canonical form and economy', p. 284-90) touches upon problems of efficiency in examining the more or less large quantity of syllables permitted by the phonological system of the languages which have been left 'uninhabited'. He advances a hypothesis: 'With a given percentage-wise utilization, we should be inclined to say that a heavy piling-up of morphs in just one portion of the used shapes is less efficient than a more even spread. Little work has been done along that line, but probably the edge of greater efficiency which Mandarin seems to show over English would be somewhat diminished if we could allow accurately for these factors.' He then defends himself immediately against the suspicion that one could find there—*horresco referens*—a judgement of value: 'It must be emphasized that the measurement or estimate of the morphophonemic economy of a language is not a value judgement. We cannot assert that greater economy is "better" in any logical, ethical, or esthetic sense. It is possible, indeed, that too efficient an economy, with some phoneme systems, might impair communication. Morphophonemic economy is simply one of the ways in which languages can differ.' (p. 289-90.) It will be observed that, after all, this question could probably be solved by a rather commonplace scientific method—by experiment; the 'possibility' that a 'too efficient economy . . . might jeopardize communication' should be susceptible of verification by adequate experiments with various different languages from this point of view. One sometimes gets the impression that Hockett, a specialist in Chinese, wishes to avoid at all costs any trend of thought which might lead him to declare

this language to be superior in some respect to his own; there is, perhaps, in him a certain nationalistic tendency, more or less unconscious or latent.

154

112. J. WHATMOUGH, *Language: a modern synthesis*, op. cit. For example, p. 45, where he states that the grammatical system of a language 'is apt to be more complex the more "primitive" the culture of the society that uses it. Most of the societies which share twentieth-century Western mechanized civilization, on the other hand, have drastically reduced their grammatical apparatus to mere remnants of irregularities. This is inevitable, for the simpler the language instrument—within limits—the more effective it is' (however, he does not define within which limits). A little farther (p. 46), Whatmough writes that the 'derivational device tends in general to be more concrete than the relational' and that the 'relational process either favors the growth of inflection . . . or, going still further, favors analysis'; he characterizes the Latin inflectional system as 'inadequate' (p. 47); regarding the *feminarum honorum* type: 'a redundancy which English has long since abandoned' (p. 127, see also a passage in Appendix 4, p. 219), on the English language which, 'giving up formative elements in favor of separate words . . . has freed itself from troublesome appendages', and another on the process of 'reshaping in adherence to a pattern' (p. 211). He puts forward a hypothesis (p. 17): 'Does the emergence of an abundance of grammatical forms, as in early Indo-European languages, the age of inflection, represent a suddenly widened understanding, which was making unprecedented efforts to cope with previously guessed at, but very involved, relationships, which their modern representatives now face with simpler linguistic devices? And the almost complete absence of variety in grammatical forms, helped out by ample gesture to indicate even fairly simple relationships, as in the Aranta of Australia, does this represent a feral stage in which not much is found to talk about?' This idea on the 'age of inflection' is encountered, in practically the same terms, but this time in the form of an assertion (p. 163). Concerning the problem of morphophonemic economy mentioned by Hockett, Whatmough takes a position which depends implicitly on the notion of redundancy brought out in the theory of information (p. 115, and see Appendix 2, p. 215): 'There is . . . a somewhat low degree of efficiency in the use of the structural resources of a language, imposed by the requirements of comprehension and communication'—but he does not thus explain the considerable differences presented, in this connexion, by the different languages (he does not, in fact, consider any but European languages, of course, of a rather uniform type). Whatmough believes, apparently, that present-day languages can be improved: 'this . . . instrument is capable of far greater refinement and more orderly application yet' (p. 158); 'language, in our time, at least, is working badly everywhere except in scientific discourse, and even there the scientists themselves express their own misgivings from time to time. . . . Improvement

of the technique can be hastened by deliberate study . . . we need not await the slow adaptation such as has taken place in historic linguistic changes' (p. 211-2). He writes: 'A finer synthesis, a sharpening and refinement, of the unique biological instrument, language, is perhaps a next possible step, by a new means of symbolization; or it may not be achieved short of some new evolutionary mutation' (p. 83). But his book ends, however, on a note of impotence: 'We do not know how our language will be redesigned—that is, amended or corrected; redesigned it must be, and if left to itself it will redesign itself . . .' (p. 213), and, in the Appendices (p. 219) only a hazardous anticipation is found ('If electronic symbolization, dispensing entirely with linguistic patterns and substituting its own, leads to a satisfactory interlingua . . .') or a prediction of detail (p. 223) on the disappearance of strong English verbs toward the year 3000.

- 154 113. B. J. WHORF, *Language, thought, and reality*, op. cit., p. 56, p. 80, p. 85, p. 140-8, p. 242-3 (see Note 44 above).
- 154 114. *Ibid.*, p. 82, concerning 'Basic English': 'English is anything but simple—it is a bafflingly complex organization, abounding in covert classes, cryptotypes, taxemes of selection, taxemes of order, significant stress patterns and intonation, patterns of considerable intricacy.' A little later on, one finds in his work an idea which comes close to that of Hockett on the constant global difficulty of languages: 'It may turn out that the simpler a language becomes overtly, the more it becomes dependent upon cryptotypes and other covert formations, the more it conceals unconscious presuppositions, and the more its lexations become variable and undefinable' (p. 183)—but it is presented as a mere hypothesis. It may therefore be appropriate to cite here the conclusion of his last writings: 'Science . . . has not yet freed itself from the illusory necessities of common logic which are only at bottom necessities of grammatical pattern in Western Aryan grammar; necessities for substances which are only necessities for substantives in certain sentence positions, necessities for forces, attractions, etc., which are only necessities for verbs in certain other positions, and so on. Science, if it survives the impending darkness (written in 1942), will next take up the consideration of linguistic principles and divest itself of these illusory linguistic necessities, too long held to be the substance of Reason itself' (p. 269-70); and see his statement (p. 240) concerning a 'contrastive linguistics' designed to permit the creation of a 'new technology of language and thought'.
- 154 115. Joseph H. GREENBERG, 'Language and evolutionary theory', *Essays in linguistics*, op. cit., p. 56-65. He revives the ideas expressed by O. JESPERSEN in his last work *Efficiency in linguistic change*, Copenhagen, 1941, by observing that he was wrong in taking an 'internally conditioned drift' toward morphological simplicity in Indo-European languages for a 'universal linguistic trend', that morphological simplicity is but an aspect of the problem—which can be accompanied by a great semantic

complexity—but that, at least, it is a ‘minor aspect of efficiency’. However, it would be necessary to measure the ‘over-all degree of efficiency’ in relation to the basic function of the language: communication (p. 61). Greenberg insists, particularly, on ‘the abolition of vagueness and ambiguity’, which could, nevertheless, result in the probable elimination of humour and poetry (p. 64). From the morphological point of view, ‘in this limited aspect the despised pidgin languages are more advanced than such cherished forms of speech as classical Sanskrit. But ‘a comparative measure of efficiency which includes all relevant phonological, grammatical, and semantic aspects had never been worked out, and, in view of the complexity of each aspect and the disparity among them, it does not appear very likely that one can be developed’—this second affirmation is very debatable; difficulty does not necessarily mean impossibility. In conclusion, he holds as certain that ‘the evolution of language as such has never been demonstrated, and the inherent equality of all languages must be maintained on present evidence’ (a note adds: ‘they are all “created equal”’ (p. 65). He comes close here, one sees, to Hockett’s position. In a later chapter, ‘Structure and function in language’, Greenberg returns to the concept of ‘functional efficiency; he submits in principle that the latter can be determined by mathematical methods (p. 81). Concerning the phonological system, he writes first that ‘That system is most efficient in which all the combinations of features are utilized’ (p. 81), but immediately thereafter he observes that a system efficient in this aspect ‘may be disfunctional when considered on the higher structural level of complete utterances, where utilization of every recourse of features would lead to insufficient redundancy’ and adds, ‘We might hypothesize, therefore, that some middle value would appear in languages as a compromise between the two functional requirements’ (p. 82). This ‘some middle value’ is, obviously, very vague. From the morphological standpoint, he mentions as a problem to be studied the ‘degree of tolerance of morphological irregularities, which can be measured by the proportion of regular and irregular constructions in samples of texts (a method which he had advocated under the name of ‘agglutination index’ in the volume edited by R. F. SPENCER, *Methods and perspectives in anthropology*, (Minneapolis, 1954); he indicates that ‘it would be possible to admit an increase of redundancy function’ for the irregularities (same idea in Siertsema, op. cit., 1955, p. 223). Likewise, the length of morphemes could be studied, of which an ‘excess of either shortness or length seems to be avoided. In semantics (i.e. in lexicology) the degree of tolerance of language as regards homonyms, the ‘extent of patterning’ could be studied (p. 83). In his last chapter, already cited many times, Greenberg proposes a hypothesis in terms of information theory on the preference of most languages for suffixing, taking into account the fact that the root offers more information than the affix (p. 91). All of the above—even though it does not indicate many results, nor

- even hypotheses which could be accepted straight away—at least offers the merit of indicating a line of approach: that of computation on specific and precise points, which offers an escape from vague generalities and entrance into the field of application of scientific methods (and which—although Greenberg, error excepted, makes no mention of it—would lend itself to controlled experimentation).
- 154 116. J. VENDRYES, 'Le langage et la vie mentale', *Conférences de l'Institut de Linguistique*, XI, p. 18-9.
- 154 117. In the discussion at the Société de Psychologie on the subject of Meillet's communication 'Le caractère concret du mot', see *LHLG*, II, p. 19-20.
- 154 118. M. COHEN, *Pour une sociologie . . .*, op. cit., it is symptomatic that the passage on 'Progrès des civilisations et transformations (progrès?) dans les langues' (p. 138) contains a question mark. In the 1947 article mentioned earlier, 'Faits linguistiques et faits de pensée', Cohen still remains rather faithful to Meillet's thesis (p. 28-9 of *Cinquante années de recherches*), suggesting even a 'connexion between . . . the division of labour and the extreme diversity of tools with a preference for analytical constructions', and the same thought appears in the 1948 brochure (*Linguistique et matérialisme dialectique* reproduced in same volume, see p. 51-2) where, moreover, an interesting suggestion is found: 'The morphological systems are in uneven equilibrium and in danger of unbalance for the general reason that they are not rational constructions.' One year later, Cohen (pamphlet *L'évolution des langues et des écritures*, 1949, 1955 volume, p. 63) raises the question: '... can it be said . . . that there has been some progress?' and replies 'A distinction should be made. Since organized languages have been constituted . . . there does not seem to have been any fundamental and constitutional progress . . . the language instrument has not been substantially modified or perfected. However, there is progress . . . in the history and the use of language'. We find ourselves here very close to Greenberg's or even Hockett's theses. In a conversation we had with him, Cohen insisted on the idea of *compensation* (e.g. the complicated spelling of the English language). But is this really a question of 'compensation', or more simply of 'unequal degrees of development' of such and such a part of the world linguistic system? See the recent study by V. TAULI, 'Standards of efficient language', *Verba docent, Juhlâkirja Lauri Hakulisen 60 vuotispäiväski*, p. 360-71, 1959, a chapter of a book in preparation, 'Foundations of practical linguistics; prolegomena to a theory of language planning'.
- 154 119. R. L'HERMITTE, 'The problems of internal laws of development of language and Soviet linguistics', In: André MARTINET and U. WEINREICH (eds.), *Linguistics today*, (New York 1954; *Word*, vol. 10, no. 2-3); he cites, particularly, (p. 71) a study in this direction by V. A. Zvegintsev. One should not, of course,

be surprised to find Russian linguists contesting a theory which would lead to considering their language as 'archaic' . . .

- 155 120. See the detailed passage by V. N. TOPOROV in his essay on probabilistic methods for the study of language, translated by the US Joint Publications Research Service, JPRS 3300, p. 9-10.
- 155 121. See the *Proceedings* of the Congress of Linguists, question C1, WHATMOUGH's report, p. 137-41; contributions, p. 142-5; discussion, p. 463-85 (notably SZEMERÉNY, p. 481-3). Question C5, contribution, p. 173-7, discussion, p. 523-41.
- 155 122. Alfred MASTER, *ibid.*, p. 539-41.
- 155 123. The best general outline remains that of MEILLET 'Différenciation et unification dans les langues', *LHLG*, I, 1911, p. 110-29; among others, the following formula will be noted, 'differentiations tend to render a language intelligible only to narrower and narrower social groups. They run counter, therefore, to the main object of language which is to facilitate relations between men' (p. 116). Concerning 'resistance to innovation', which is strong among the great common languages of civilization, 'because innovation must extend to a great number of subjects distributed over a very large geographical area. And this is a great blessing . . . It is the role of the school and of literature to maintain linguistic unity once created' (p. 121). Meillet, always preoccupied with discovering general laws, believed he had found another one here: 'progress in linguistics tends to bring this dominant factor in the history of languages more and more to the fore: the creation and extension of these common languages, which are the product of the unity of civilization, covering greater or lesser fields. . . . Sooner or later, and at times immediately, each differentiation is followed by a reaction which tends to re-establish or to create a unity of language wherever there is unity of civilization' (p. 129). It is therefore not surprising that, as observed by COHEN (*Pour une sociologie* . . ., p. 334; there is an excellent chapter in this work, p. 307-35 concerning these questions) he conceived some 'irritation at the large number of languages being accepted as cultural languages in the Soviet Union', and that he expressed it in the second edition of *Les langues dans l'Europe nouvelle*, Paris, 1928. He had not perceived any other causes for the multiplication of cultural languages, despite a 'common civilization', which have acted in full during the past forty years with the diffusion of education in countries until now largely illiterate, a diffusion which proceeds necessarily, through the promotion of the 'national' language to the rank of cultural language (see, on this subject, J. STALIN, *Marxism and the national and colonial question*, *passim*). The article by Lo Tchang-Pei and Lu Chou-Sang mentioned earlier (p. 150) is an important contribution to the present extension of a common language; another interesting article by Ajoy Gosh in the same journal (p. 140-61) deals with the much more complex problem of a common language in India. A further observation by Meillet

(*LHLG*, II, p. 78) on the rapid change in Latin and its profound differentiation between the third and ninth centuries: 'inasmuch as culture declined considerably and relationships between populations became difficult and relatively rare, the language changed, encountering no resistance in the conservatism of the cultured elements and in the necessity of maintaining unity of language throughout the entire field.' He emphasizes, on the contrary, the stability of classical Latin, as of the French language, since the seventeenth century (*ibid.*, I, p. 121).

Hockett (*A course in modern linguistics*, op. cit., p. 367-8) takes a completely negative position on the question of the maintenance of the stability of a language. He writes that 'conscious efforts to impede the natural slow change of speech habits have always failed. For example, the French and Spanish Academies have succeeded, in the last few centuries, only in temporarily fixing certain minor habits of spelling and of the formal style of the language used in writing'. See, for French, certain very cogent observations—which run counter to this thesis of Hockett—in WARTBURG, op. cit., p. 24-5, p. 177; M. COHEN, *Notes de méthode pour l'histoire du français*, Moscow, 1958, p. 59-65; etc. Naturally, this does not reflect merely (nor even especially) the influence of the Académie Française.

Whatmough notes (p. 56) that 'if the ease and certainty of modern telecommunications remain even as effective as they are now, or still more if they are improved and perfected, then the forces which in the past have led to the dissolution of a common language into manifold dialects, and these in their turn into separate, national languages, will be greatly mitigated and perhaps cease altogether'.

- 155 124. 'To the inventor of an artificial language, the process of invention may seem like a manifestation of his own free will; looked at from outside, however, we see that the inventor's decisions are based on his own of speech, his knowledge, accurate or inaccurate, of various other languages, and his general understanding or misunderstanding of how language works. It is no accident that most of the last hundred years' crop of artificial language, including Esperanto, are clearly classified as European languages in their semantics, their grammar, and their phonology. If an investigation of Esperanto were carried on by specialists who were ignorant of its origin, they would class it as an aberrant form of Romance. They might even venture the guess that the features which render it aberrant were due to pidginization. This guess would not be far wrong, for the process of inventing actually involves a continual borrowing of forms and features from this, that or the other natural language, with irregular and unpredictable distortions of shape' (HOCKETT, op. cit., p. 422-3).
- 155 125. M. COHEN (*Pour une sociologie du langage*, op. cit., p. 349-50) blames JESPERSEN (*Language*, op. cit., p. 216-36, containing a bibliography to 1918) for failure to distinguish sufficiently the pidgins and the creoles 'which in structure have some points in

common, but different origins and development', he discusses the pidgins on p. 338-9 and p. 349-50 and the creoles on p. 285 and p. 290. To his bibliography should be added: Robert A. HALL, Jr. *Melanesian pidgin English*, Baltimore, 1943, and *Hattian creole*, American Folklore Society (Report 43), and E. JOURDAIN, *Du français aux parlers créoles*, Paris, Klincksieck, 1956. The fundamental difference, as indicated by Hockett, is that the creole dialects are taught to children as mother tongues, whereas the pidgins are second languages of adults. F. BODMER (*The loom of language*, London, Allen & Unwin; New York, Norton, 1944) notes (p. 447) that apart from Steiner none of the 'pioneers of language planning' seem, indeed, to have considered the pidgin or the creole languages as deserving of sympathetic study.

- 155 126. The most complete historical outline, up to 1906, is that found in L. COUTURAT and L. LÉAU, *Histoire de la langue universelle*, Paris, Hachette, 1903, with a supplement under the title *Les nouvelles langues internationales*, 1907, but this work is far from impartial (the authors were the creators of Ido). Albert L. GUÉRARD, *A short history of the international language movement*, (New York, 1922), is more objective, but very brief; we have been unable to secure a copy of the work, written in Esperanto, by the Russian E. DREZEN, *Historio de la mondalingvo*, Leipzig, Ekrelo, 1931. A bibliography up to 1928 is found in P. E. STOJAN's *Bibliografio de internacia lingvo* [Bibliography of international language], Geneva, 1929. Brief outlines in M. COHEN's *Pour une sociologie du langage* (p. 342-3 and p. 351-3) and CARROLL, *op. cit.*, p. 125-32 (especially on the subject of Basic English); in J. G. HOLMSTROM (ed.), *Scientific and technical translating and other aspects of the language problem*, Paris, Unesco, 1957, 2nd ed., 1958; there is a descriptive chapter and summary of various opinions (chapter 6, p. 173-208). The book by H. JACOB, *A planned international language*, London, Dennis Dobson, 1947, is somewhat disappointing; on the other hand chapters XI and XII of Frederick BODMER's *The loom of language*, *op. cit.*, p. 448-518 of the American edition, are very interesting, even though all the conclusions are not accepted. We have devoted a chapter of our *Théorie et pratique des classifications documentaires*, Paris, UFOD, 1956, p. 84-96, to relationships between classification and constructed international languages, but it requires serious revision and amendment.
- 155 127. See above Note 86 and 'The function of an international language', *Psyche*, 11, 1931, p. 4-15, reproduced in H. N. SHENTON, E. SAPIR, and O. JESPERSEN, *International communication*, London, 1931, p. 65-94, and in *Selected Writings of Edward Sapir*, 1949, p. 110-21, which contain a criticism of the inconsistencies and difficulties of English (p. 114-5) and of French (p. 116), as well as a research programme towards a more satisfactory international language (p. 119); we have not seen Sapir's 'Memorandum on the problem of an international auxiliary language', *Romanic Review*, 16, 1925, p. 244-56.

- 156 128. O. JESPERSEN, *Eine internationale Sprache* [An international language], Heidelberg, 1928.
- 156 129. A. MARTINET, *Questionnaire rédigé au nom d'IALA*, New York, 1946 (prepared in co-operation with J. P. Vinay); see also his report to the sixth International Congress of Linguists, 1948, *Actes*, Paris, 1949, p. 93-112 and p. 585-600.
- 156 130. J. WHATMOUGH, *Language*, p. 58-65; especially concerning ILA (International Language for Aviation), p. 59; a critique of the IALA Interlingua, p. 61; debatable ideas on English as an international language, p. 62 and p. 64. Whatmough adds a note: 'In comparing dialects correlation methods may be used to show how much the bilingualism of frontier districts, or the polylingualism of an interlingua, may safely draw from different languages. The inventors of artificial interlinguas seem to be unaware of this important procedure.' (p. 214-5.) This observation can apply to 'interlinguas' of the Interlingua type, which seek merely 'a smaller common denominator' among a group of closely allied languages; it is doubtful whether it can serve in the elaboration of a truly international interlingua (not linked to any particular linguistic group).
- 156 131. The most interesting by far having been the great mathematician and logician Giuseppe PEANO of Turin, promotor of the *Latine sine flexione* [Latin without inflexions]; see F. Bodmer, *The loom of language*, op. cit., p. 471-6. Louis COUTURAT, who 'rediscovered' the works of Leibniz on international language (see *La logique de Leibniz*, Paris, Alcan, 1901, a basic work which, disgracefully, has not been republished), wrote several articles: 'D'une application de la logique au problème de la langue internationale', *R. Métaphys. Morale*, 1908, 761-9; 'Du rapport de la logique et de la linguistique dans le problème de la langue internationale', *ibid.*, 1911, p. 509-16; 'Sur la structure logique du langage', *ibid.*, 1912, p. 1-24; 'Pour la logique du langage', *Bulletin de la Société française Philos.*, 1913, p. 135-65; see the discussion at the SFP, *ibid.*, 1912, p. 47-84. Here again, as in his *Histoire*, Couturat especially advocates Ido.
- 156 132. B. C. VICKERY, 'The significance of John Wilkins in the history of bibliographical classification', *Libri*, 2, 1953, p. 326-43, which, however, does not envisage in detail the strictly linguistic viewpoint.
- 156 133. STUART C. DODDS, 'Tilp: a ten-letter alphabet of meanings', *General Semantics Bull.*, nos. 6-7, 1951, p. 38-44; *An alphabet of meanings*, Washington Public Opinion Laboratory, Seattle (multilithed document U-56-107, 1956); the first presentation was in 1947 in *Systematic Social Science*, p. 679-85, Seattle, University Book Store. See our sixth report to the FID/CA Committee, September 1958 (multilithed document entitled 'Les progrès et l'avenir du "langage classificatoire"', p. 10-11). 'Tilp' would probably have pleased Whorf: it is, indeed, a language of a most highly 'oligosynthetic' type (in accordance

with the term created by Whorf, see his 1956 work, p. 12-13; p. 392 of Hoijer's work, *Linguistic structures of native America*; and some as yet unpublished texts such as the 'Notes on the oligosynthetic comparison of Nahuatl and Piman', 1928), Dodd is intent on reducing the 'connective' and 'operative' parts of language to combinations of 10 elements, distributed among 3 classes: qualitative, quantitative, and relative—in line with Kant's Categories of the Understanding which the author has reduced to mathematical expressions' ('Tilp: a ten-letter alphabet of meanings', *ibid.*, p. 38). The order of elements (prefixes or suffixes) plays a great role in the system (*ibid.*, p. 39). Dodd used the 16 operators and the 84 connectives of Basic English, which he translates into various combinations of his 10 basic letters. (*ibid.*, p. 40) Also to be noted is the 'quantic classification of grammatical parts of speech' (*ibid.*, p. 42), which, moreover, is entirely artificial.

- 156 134. Duplicated document, San Diego, California, 1953. See on this subject our seventh report to the FID/CA Committee (duplicated, 1960), p. 30. The 'Lincos' of H. FREUDENTHAL (*Lincos: design of a language for cosmic intercourse*, vol. 1, Amsterdam, North Holland, 1960) is an attempt at a 'logisticized' language. The 'Loglan' of James COOKE BROWN ('Loglan', *Scientific American*, 202(6), 1960, p. 53-63) is a rather curious mixture of words 'drawn' according to a 'learnability score' from eight widely diffused languages and of 120 'connectors' 'operators', 'indicators' and 'phrase operators' made up 'a priori'.
- 156 135. For example Paul MITROVITCH, *An attempt at an inter-systemal grammar of auxiliary languages*, Sarajevo, 1953; *An essay on interlinguism*, Sarajevo, 1953; *Les problèmes interlinguistiques*, 2nd ed., 1940; Federigo Beigbeder ATIENZA, 'La normalización internacional de la nomenclatura y simbología técnica' [International standardization of technical nomenclature and symbolism], *Revista del Instituto nacional de racionalización del trabajo*, 11, 1958, p. 513-23.
- 156 136. It is doubtless on the latter point that a comparative study of these different systems would be of most interest; the elements on this subject are in H. Jacob's book already cited; See also: P. MITROVITCH, 'Word formation', *An attempt . . .*, op. cit., p. 13-9; Beigbeder Atienza's article; the work of the ISA 37 Committee in 1934-37, especially the report by DREZEN, CHATELAIN, SPIELREIN, SHIRKOFF, *Ueber einen internationalen terminologischen Code* [On an international terminology code], Commission for Technical terminology of the Academy of Sciences of the USSR, November 1935, English version, June 1935; E. WÜSTER, *Konturoj de la lingvo-normigo en la tekniko* [Outline of normal language in technology], Budapest, Literature Mundo, 1936. In Esperanto, a curious element, not without interest, is the table of particles, in double-entry; indefinite, interrogative, demonstrative, universal, negative, for the columns; quality, motive, time, location, manner, possession, thing, quantity,

individuality, for the rows. Concerning the IALA Interlingua, see the 1945 *General report*, the Interlingua-English dictionary and the grammar prepared under the direction of Alexander GODE (New York, Storm Publishers, 1951), and, by the latter, a pamphlet *The case for Interlingua*, reprinted from the *Scientific Monthly*, August 1953, in which Gode refers very abusively to Whorf in order to bolster up his thesis.

- 156 137. Elimination of articles, of all final grammatical vowels (except in the case of the third person of the personal pronoun), of the compulsory plural, of the inflexion of pronouns, of all verbal affixes except those of participles, of compulsory distinctions of tense. Peano attaches only very little importance to rules which should govern the order of words in such a language; he is not interested in phonological aspects, nor in the problem of vocabulary reduction by a combination of simple terms.
- 156 138. The two best abridged critical outlines are those of CARROLL (op. cit., p. 128-32) and BODMER (op. cit., p. 479-84); the first provides the essentials of the bibliography. We have not seen I. PALDAREV'S study (*Philologica*, 6, 1951, p. 29) mentioned by COHEN (*Pour une sociologie du langage*, p. 353). Concerning the special verbal system of Basic English, see the criticisms of Carroll (p. 130) and Bodmer (p. 483-4). Ogden had created, in order to spread his ideas, the Orthological Institute, with a principal office in Cambridge (England) and a certain number of branches. The institute published the review *Psyche*, containing numerous interesting studies, outside and beyond Basic English itself. We shall not discuss here 'Français élémentaire', which does not relate directly to our subject. The most fundamental objection to Basic English comes from Whorf (op. cit., p. 82-3). After having reviewed all of the structural complexities of the English language, in the terms quoted above (Note 114), he adds: 'As with Basic English, so with other artificial languages [he refers here to the 'naturalistic' languages], underlying structures and categories of a few culturally predominant European tongues are taken for granted; their complex web of presuppositions is made the basis of a false simplicity.'
- 157 139. Lancelot HOGBEN, *Interglossa: a draft of an auxiliary for a democratic world order, being an attempt to apply semantic principles to language design*, Harmondsworth, Penguin Books, 1943. E. Buysens was certainly right in drawing the attention of the seventh Congress of Linguists to Hogben's work (p. 22 of the *Proceedings*); this small book will certainly be epoch-making—which does not imply that all of the ideas contained in it should be endorsed. In particular, the criticisms made of the auxiliaries of Basic, apply in large part to Interglossa's 'system of operators' (see chapter III, p. 42-55). His basic principle, that the vocabulary of an international auxiliary language must be based on 'internationally current Greek roots' (or sometimes Latin), for the latter are the 'truly international roots' (p. 12-14 of chapter 1, 'Interglossa and its predecessors', which should

be read in its entirety) is at the very least debatable: 600 million Chinese (among others) do not use them. Hogben lifted from Esperanto the method of indicating each main class of words by a special vowel ending (see p. 37); this is not much better done by Hogben than by Zamenhof.

- 157 140. *Zsady powszechnej ideografiki analitycznej. Prinzipien der allgemeinen analytischen Ideographie* [Principles of general analytical ideography], Krakow, Prace komisji jezykowej Polskiej Akademji [Works of the Commission on Linguistics of the Polish Academy], 1925.
- 157 141. See the Unesco report, edited by Holmstrom, cited earlier (Note 126), p. 203-5.
- 157 142. *Semantography*, Sydney, Bliss Institute, 1946-49. (Bliss Institute, 2 Vicar Street, Googee-Sydney, N.S.W., Australia.)
- 157 143. Holmstrom, op. cit., mentions it merely (p. 203, in 11 lines) on the basis of a second-hand account by W. A. Heaney, published at Yellow Springs, Ohio, in 1952.
- 157 144. We mentioned it briefly in *Les progrès et l'avenir du langage classificatoire*, 6th report to the FID/CA Committee, 1958, p. 12-13. We refer the reader especially, on the subject of the present report, to the chapter entitled 'Grammar, or the rules of the game', p. 124-51 (three categories, indicated by superior symbols, which are, however, optional: object, action, value); p. 183-99 (tense, active, and passive), and p. 243-6 (expression of English -able and -ed); the chapter 'Anarchy in language', p. 288-318 (particles); finally 'The future of words', p. 423-33 (composition, derivation). One may mention also, as derived from this same ideographic trend, Jean François Rozan's interesting little book, *La prise de notes en interprétation consécutive*, Geneva, Georg, 1956, which described (p. 13-23) a technique for noting ideas expressed in speeches by means of seven principles (stressing particularly the necessity for respecting the 'logical sequences') and (p. 28-35) 20 basic graphic symbols. We have not seen the universal language system devised by M. Heimer, *Mondial* (Lund, 1947), nor Sidney E. CULBERT's *Remarks on the semantic structures of certain proposed interlanguages* (report in *Psychol. Abstracts*, 30, 1956, p. 246); we regret, therefore, being unable to discuss these here.

Problems of symbolization and notation

- 159 1. *Théorie et pratique des classifications documentaires*, p. 22-32, 297-301 and *passim*, 1956, see the index, under 'symbolization'; see also our seven reports to the FID/CA Committee, from 1953 to 1958.
- 159 2. See the reports on the work of the Committee on Linguistic Statistics of the Permanent International Committee of Linguists, initiated by B. TRNKA, *A tentative bibliography*, Utrecht, Spectrum Publ., 1950; G. HERDAN, *Language as choice and*

chance, Groningen, Noordhoff, 1956, a very debatable book, especially in the general conclusions which he believes can be drawn from an examination of individual facts, but containing much interesting data; by the author, 'The numerical expression of selective variation in the vowel-consonant sequence in English and Russian', *Studies ... Whatmough*, 91-104, 1957; V. BELEVITCH, 'On the statistical laws of linguistic distributions', *Annales Soc. Scientifique Bruxelles*, 73, 1959, p. 310-26. As indicated, in particular, in this last work, one is now led to revise certain previous hypotheses (such as the Estoup-Zipf-Mandelbrot Law). Belevitch discussed, at the GRISA Seminar of Euratom, on 17 February 1960, 'problematical mathematical linguistics'; the same day, P. Braffort presented at this seminar his 'certain mathematical linguistics'. J. BERRY in 'Some statistical aspects of conversational speech', In: Willis Jackson (ed.), *Communication theory*, p. 392-401, London, Butterworth, 1953, demonstrated that it was necessary to take into consideration, in studies of frequency of phonemes, their accent and their location ('details' to which heretofore insufficient attention had been given).

- 159 3. See, among other recent studies bearing on the subject of this report, Sol SAPORTA's 'Frequency of consonant clusters', *Language*, 31, 1955, p. 25-30; Einar HANSEN's 'The syllable in linguistic description', *For Roman Jakobson*, p. 213-21; W. FUCKS, 'Die mathematischen Gesetze der Bildung von Sprach-elementen aus ihren Bestandteilen' [The mathematical laws of the formation of linguistic elements from their components], *Nach. tech. Fachber.*, 3, 1956, 7-21; H. M. MOSER and J. J. DREHER, 'Evaluation of the military alphabets', *Speech monographs*, 22, 1955, p. 256-65; G. A. MILLER, 'The perception of speech', *For Roman Jakobson*, p. 353-60, op. cit.; P. E. NICELY, 'An analysis of perceptual confusion among some English consonants', *Journal of the Acoustical Society of America*, 27, 1955, p. 338-52; A. M. LIBERMAN, 'Some results of research on speech perception', *ibid.*, 29, 1957, p. 117-23. We have not seen Stanley M. SAPON and Ezra V. SAUL, 'Findings on the differential resistance to noise of French, Spanish, and English', mentioned by J. B. Carroll in *Psychological Abstracts*, 30, 1956, p. 247.
- 159 4. See Otto JESPERSEN, *Language*, p. 396-11, op. cit.; Edward SAPIR, 'A study of phonetic symbolism', *Journal of Experimental Psychology*, 12, 1929, p. 225-39, reproduced in *Selected Writings*, 1949, p. 61-72; Allan WALKER READ, *English words with constituent elements having independent semantic value*, Philadelphia (The Malone anniversary studies, edited by T. A. Kirby and H. B. Woolf, Baltimore, 1949); Roger W. BROWN, Abraham H. BLACK, Arnold E. HOROWICZ, 'Phonetic symbolism in natural languages', report by N. L. SOLOMON in *Psychol. Abstr.*, 30, 1956, p. 246; A. DANIELOU, 'L'alphabet sanscrit et la langue universelle', *Lotus bleu*, 61, 1956, p. 51-68; Swami NIKHILANANDA,

'Aum: the word of words', In: Ruth N. ANSHEN (ed.), *Language: an enquiry into its meaning and function*, p. 80-5, New York, Harper, 1957.

- 159 5. For the latter, see N. TRUBECKOR's article 'Wie soll das Lautsystem einer künstlichen internationalen Hilfssprache beschaffen sein' [How should the phonetic system of an artificial international auxiliary language be organized], *Travaux Cercle ling. Prague*, 8, 1939, p. 5-21. We again draw attention to Z. DOBROWOLSKI's very interesting work on the symbolization of documentary coding *Budowa klasyfikacji* [Structure of classification] Warsaw, Państwowe Wydawnictwa Techniczne, of which a French translation is in preparation.

