UNIVERSAL CLASSIFICATION I
UNIVERSAL CLASSIFICATION I
Subject Analysis and Ordering Systems.

Proceedings
4th Internat. Study Conference on Classification Research,
6th Annual Conference of Gesellschaft für Klassifikation e.V.

Editor: INGETRAUT DAHLBERG
Gesellschaft für Klassifikation e.V.
International Federation for Documentation,
Committee on Classification Research, FID/CR

Frankfurt INDEKS VERLAG 1982
INTRODUCTION

This volume contains those papers of the 4th International Study Conference on Classification Research and 6th Annual Conference of the German Society for Classification which were delivered by their authors for preprinting ahead of the conference. They are arranged in the sequence of the contents list following the program of the conference. The papers still missing have been given a star (*). They will be printed in vol. II after the event together with discussions, reports on the sessions, introductory words of the sponsors of this conference as well as with the conclusions and recommendations. Abstracts and indexes to the two volumes will also be included in vol. II.

Volume I starts out with a review of the 3rd International Study Conference on Classification Research of FID/CR which took place at Bombay, Jan. 6-11, 1975. The review writer is a university librarian from Jena, GDR, our colleague Dr. Helmut B e c k, well known for his many articles on classification, mostly published in German in the Zentralblatt für Bibliothekswesen. We are very grateful for his attempt to link the two conferences together and for his endeavours to pinpoint the essential results of the last conference in order that these may not be forgotten during the forthcoming deliberations.

Since it was not possible to do any major editing work on the papers after delivery in camera-ready form we beg our readers' pardon for any errors and mistakes encountered. We owe many thanks to our colleague, Prof.Dr.Martin S c h e l e and his printing house through whose good services the volume could be printed in only two weeks' time, just before the conference date.

May the efforts of our colleagues to offer their ideas and to write them up in time and in the form requested (including observance of a limitation of 8 pages only for each paper) render the necessary service to all participants of the Augsburg Classification Conference in order that its results will push the world's classification knowledge forward for the benefit of all those dealing with the problems of the organization, identification and retrieval of knowledge.

Ingetraut Dahlberg, Chairperson
FID/Committee on Classification Research
Gesellschaft für Klassifikation e.V.

June 12, 1982
CONTENTS

Introduction
H. BECK: Book Review of the Proceedings of the 3rd Intern. Study Conf. on Classification Research, Bombay, Jan. 6-11, 1975 3

Topic 1: General Principles and Policies of Classification Systems

A1 E. de GROLIER: Classifications as a cultural artefacts 19
B1* E. OESEKR: The two systems for the organization of knowledge: On the foundation and on features of a universal background system 35
C1 D.J. FOSKETT, S. BURY: Concept organisation and universal classification systems 35

la1 P. ROLLAND-THOMAS: Towards the establishment of the validity of encyclopedic library classification systems 44
la2* C. CAVALCANTI: System applications and concepts in the field of classification 44
la3 C.M.M. KASHYAP: General theory of measurement and theory of subjects 51
la4 H.H. WELLSCH: A new classification of the world’s languages 59
la5 J. MILLS: Practice and theory in a general classification: the new BLISS Classification (BC2) 69
la6 M.A. GOPINATH: Application of General Systems Theory to the classification theory of Ranganathan 78
la7* P.N. KAULA: Rethinking on concepts in the study of classification 86
lc1 D. AUSTIN: Basic concept classes and primitive relations 86
lc2a H. FELBER, W. NEDOBITZ: The basis of Wüster’s classification research 95
lc2b W. EISENWORT: Remarks on certain foundational problems of a general theory of terminology 102
lc3 W. GÖDERT: Library subject analysis in tension between universality and subject speciality 106
lc4 E. WAHLIN: Analysis of different ordering principles in the area ‘Technology’ (T), ‘Industry’ (I) and ‘Material Culture’ (M) - the ELK-Project 114
lc5* P.E.V.d. VET: Notes on the foundations of classification theories and Klassifikation

lc6* H. MÖNKE: Analytische Pragmatik versus Dogmatik in der Klassifikation

Topic 2: Structure and Logic of Indexing Languages

A2* E. FUGMANN: Natural language versus indexing language in mechanised information retrieval 125
B2 G. BHATTACHARYYA: Classaurus: Its fundamentals, design and use 139
C2 T. HENRIKSEN: On the mechanisation of the chain index 139
2a1 Ph. RICHMOND: Classification in syntetic structures 149
2a2 T.N. RAJAN, B. GUHA, R. SATYANARAYANA: Association relationship of concepts as seen through citations and Citation Index 157
2a3* T.G. CHAYEF: The representation of facets in a general concept network for index display generation 157
2a5 G. RAHMSTORF: Explication and translation of existing classification systems 166
2a6 M. GIJANIK: Fuzzy faceted thesaurus construction 166

* Paper numbers with a star are not included in Vol.I.

2c1 F.J. DEVADASON, M.KOTHANDA RANUJAM: Computer-aided construction of an 'alphabetic' classaurus 173
2c2 G. SECHSER: Theoretical generalizations in retrieval system research, their objectivity and relevance 183
2c3 KRISHAN KUMAR: Theoretical bases for universal classification systems 190
2c4 P. GILKERT: Gradational classification: a topic-tagging scheme for computerised production of a scannable list 198
2c5 A. MARK PEITRSEN: A new approach to the classification of fiction 207
2c6 J. AUSTIN: The AMP classification system for fiction: Trial applications and retrieval tests 216
2c7 R. BONNER: Computer Information classification research project: user oriented empirical methods of classification construction 227

Topic 3: Empirical Investigations and Practical Use

A3 H. BORKO: The role of classification in online retrieval systems and automated libraries 235
B3 B. KELM, B.WAASSEN: Weiterentwicklung der Sacherschließungsarbeit an der Deutschen Bibliothek 246
C3 E. JOITA, P. ATANASIU, V. TEODORU: The use of ordering systems by different user groups 254
3a1* M. DIENES: Structural differences in classification systems and the testing of the compatibility matrix in the field of culture 260
3a2 P.A. COCHRANE: Classification as a user’s tool in online public access catalogs 269
3a3 I.L. TRAVIS: Faceted classification in an online environment 269
3a4 N.J. WILLIAMSON: Videotex information retrieval systems: the logical development and optimization of tree structures in a general online interactive system 277
3a5 A.H. SCHABAS: Videotex information systems: complements to the tree structure 285
3a6* H.G. KÖRNER: Classification and small computers 285
3c1 G. WERSIG, M. BURKART-SABSOUB: Empirical classification research as a basis of information policy 292
3c2 G. MALANGA: Classifying and screening journal literature with citation data 299
3c3 W. RIGBY: The use of the UDC in automated terminology data banks 305
3c4* E. SCHIBOR: Polish subject field classification, broad ordering system for use on a national scale 311
3c5 N. SWAYDAN: The universal classification and the needs of libraries in developing countries 321
3c6* F. KLEJZCZ: Towards a classification, unification and selection of automatic indexing methods

Topic D1: Numerical Methods in Classification

D1.1 N. MEDER: Family resemblance and empirical classification, cluster analysis: its philosophy of science and strategy of research 329
D1.2 J. KRA: Models of the measurement of similarity 337
D1.3 G. LOFFLER: Cluster-Strukturvergleiche - Ein Ansatz zur Zusammenhangsanalyse multivariat definierter Raumstrukturen 345
D1.4 E.V. KRISHNAMURTHY, H.S. VENKATESHAK: A relational data-base system and query language for storage, manipulation and retrieval of virus data 353
The libraries and library organization in developing countries should play an important role in this project, specially in two directions:

a. Allocate the necessary funds needed for the extra work of the additions and changes in the system(s).

b. ‘Selected group of librarians and scholars in different field work closely with the editors of the proposed classification scheme to give the right information on local subjects to insure accurate and complete treatment.

5. Conclusion

We know it is not an easy task to make a complete universal classification system which covers all fields of knowledge thoroughly, especially local subjects in different cultures. But it is possible to improve the existing classification systems as much as possible to meet most of the needs of libraries in the developing countries.

Librarians in the developing countries do not blame the Americans and Europeans for not making their classification schemes suitable for other nations. In fact it is the responsibility of the developing countries to prepare the needed classification schemes; but we discuss this matter from the universal point of view, especially most of the Asian and African librarians lack the experience and capability to prepare good and complete systems.

With more cooperation and understanding, it is possible to find the right solution for this matter, especially if European and American individuals and organizations show their interest in helping the Asian and African libraries by making the necessary expansion of their standard classification schemes including additions and changes to meet more needs of libraries in developing countries.

If these needs are met to entire satisfaction, the existing classification schemes will be considered universal classification schemes.

Without achieving this goal, none of them is considered a universal classification in its scope. I have put before you the problem we are facing, and I have also explained the ways to its solution. It is upto you now to decide.
by cluster analysis turn out to be typological concepts. For the definition of their contents we are only given the space of properties, in which the objects are so widely scattered that they are only connected with each other by complex relations of properties. The criticism of concepts generated by cluster analysis is the same as the criticism of typological concepts, as e. g. presented by HEMPEL (1).

First: Typological concepts are in most cases semantically ambiguous; i. e. it is impossible to specify common properties which are all valid for all the objects covered by the concept.

Second: Typological concepts have a vague margin. They are open in their application, i. e. they do not have clear-cut boundaries. This point refers to the pragmatics of typological concepts, for the openness means that there are always objects (at the "edges of the concept"), to which the predicate can be awarded or denied.

Third: Typological concepts are relational concepts. They are comparative and rather express a more-or-less than allow dichotomous decisions. Though this is a logical-syntactical momentum, the problem lies in the fact that this complicates the use of typological concepts.

Further down I will show that these so-called weaknesses or disadvantages of typological concepts can just as well be interpreted as their strength or advantages.

For clusters these three points of criticism can be reformulated as follows:

First: A cluster cannot be defined in one dimension of the "property space" (semantic ambiguity).

Second: A cluster does not have clear-cut edges, for it is situated in the "object field", in particular this refers to the questions of generalizing clusteranalytic results, i. e. their pragmatics.

Third: Clusters have an inherent quasi-order, i. e. as far as their concentration centres as their tertium comparationis are concerned the objects are related to each other in a comparative manner.

According to WITTGENSTEIN most of the expressions of our language are (typological) concepts in the sense of family resemblance. As an example WITTGENSTEIN presents the words "game" and "number". Game cannot be defined by one property only, which is common to all objects called game and distinguishes them from other objects which are not games. Though you can find a common property for two games at a time, for different pairs of games these common properties are mostly different. Therefore WITTGENSTEIN talks of similarities.

I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. And I shall say: 'games' form a family. (1, § 67)

In the following WITTGENSTEIN shows that concepts of family resemblance are - because of the fact that they are only based on a similarity relation - open and porous, without clear-cut boundaries, context-dependent, and inherently comparative. They are only defined by their use, which shows up in examples. In their application we follow a rule, without being able to name it explicitly:

And for instance the kinds of number form a family in the same way. Why do we call something a "number"? Well, perhaps because it has a - direct - relationship with several things that have hitherto been called number; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibres. (2, § 67)

3. Family resemblance as the philosophy of science for cluster analysis

The parallel to cluster analysis strikes at once. We have a concrete field of objects, which are connected by fibres, i. e. the multifarious similarity relations, and we understand the threads woven by cluster analysis at the connected objects only. In this function the objects become examples.

At this point of the PHILOSOPHICAL INVESTIGATION it is WITTGENSTEIN's intention to show a use of concepts that obviously follows rules, which however cannot be named. By describing such a use he tries to discover the "substance" of language resp. of the language game. In doing so he criticizes the narrow idea that a concept must have clear-cut boundaries.

Frege compares a concept to an area and says that an area with vague boundaries cannot be called an area at all. This presumably means that we cannot do anything with it. - But is it senseless to say: "Stand roughly there"? Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand - as if I were indicating a particular spot. And this is just how one might explain to someone what a game is. One gives examples and intends then to be taken in a particular way. - I do not, however, mean by this that he is supposed to see in those examples that common thing which I - for some reason - was unable to express; but that he is now to employ those examples in a particular way. Here giving examples is not an indirect means of explaining - in default of a better. For any general definition can be misunderstood too. The point is that this is how we play the game. (I mean the language-game with the word "game"). (2, § 71)
The refutation of the universal claim of classical concept generation which is definition per genus proximum et differential specificam, produces within the conception of family resemblance the motive to introduce a concept via an example. Thus the example constituted the concept: The case defines the type. It is just in this way that cluster analysis proceeds; it is therefore the best method to generate a typology.

Via the clusters as classes of examples we can now try to get to know the type they represent. It is our task to interpret and to bring into line the complex similarity relations between the examples. The state of affairs is the same as with family resemblance. The aspects (the constitutive properties) of the resemblances change from case to case, which is not astonishing in the face of the complexity of the cases.

Taking into consideration that clusters often present concentrations within the field and that the centres of these concentrations are figured out and supply us with the ideal case, the ideal type so to speak, it is the case that cluster analysis goes along with case analysis in the neighbourhood of centres. It is the case, not the single property that defines the intensity, the contents of the clusteranalytic concept.

It should now be clear that WITTGENSTEIN's theory of family resemblance forms the philosophical basis of cluster analysis. In his philosophy of language KUTSCHERA has presented, with a different intention though, a logical reconstruction of WITTGENSTEIN's conception, which confirms this in detail (3, pp. 190-203).

4. Refutation of the criticism

With his analysis of family resemblance WITTGENSTEIN also turns against a conception of so-called exact science that has become a dogma: One might say that the concept 'game' is a concept with blurred edges. - "But is a blurred concept a concept at all?" - Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need? (2, § 71)

For imagine having to sketch a sharply defined picture 'corresponding' to a blurred one. In the latter there is a blurred red rectangle; for it you put down a sharply defined one. Of course - several such sharply defined rectangles can be drawn to correspond to the indefinite one. - But if the colours in the original merge without a hint of any outline won't it become a hopeless task to draw a sharp picture corresponding to the blurred one? (2, § 77)

What does it mean, if it is really true that the original is indistinct, if indistinctness is just a property of the object itself? Is it not the case then that family resemblance and cluster analysis are the scientifically most exact way to comprehend their subject-matter? Do not the so-called weaknesses then turn into strength?

5. The unique efficiency of cluster analyses

Social science that proceeds according to a theory of action is confronted with an area of indistinctly contoured objects, for social action is characterized by the double contingency of ego-alter-syntheses. What does this contingency consist of? Action is selection within the horizon of alternative possibilities. Thus action is sensefully intentional. Sense just means the exclusion of alternative actions, which however are preserved as alternatives. They form a resource so to say and in principle must be looked upon as evenly-probable for action partners. Therefore actions in social contexts are always accidental, i.e. contingent. Nevertheless social action is ordered by rules; the potential alternatives of action are reduced, even though the boundaries are indistinct. One sees at once that the space of action is structured like a field into typical contexts of action. For WITTGENSTEIN this fact is the reason for identifying - under the aspect of meaning - the action space as language and contexts of action as language games.

Here the term "language-game" is meant to bring into prominence the fact that the speaking of language is part of an activity, or of a form of life. (2, § 23)

And to imagine a language means to imagine a form of life. (2, § 19)

If actions are defined as selective meanings within the language game, and if the language game, however, despite all its regularity is afflicted with principal contingency, then the researcher is forced in his description of the meanings of social actions to take this contingency into consideration resp. to describe it.

It is just the concepts of family resemblance and cluster analysis that are suitable for this task. By means of margins of vagueness they allow contingency, which by the way from a moral point of view must be interpreted as the liberty of action. If research in social sciences proceeds according to this terminology, another problem arises: If the results of the description of social reality contain its contingency, the claim for validity necessarily becomes finite, i.e. reduced to the past and to those cases that the description is based on. This renunciation of prognostic valency and of generalizability seems to make the results of such a research work particular-
ly prone to refutation, insecure, and contingent. The question is, how science can cope with such an inherent contingency, with such an insecurity. Let us look at this problem from another point of view:

Science itself is a form of social action. Like every action, scientific action, too, is selective and in neutralizing alternative possibilities a reduction of the complexity in the space of action. According to N. Luhmann the reduction of complexity is carried out in three dimensions: 1. in the social dimension, 2. in the object dimension and 3. in the time dimension (4, pp. 48-61). Reduction of complexity in the time dimension means for scientists that they obtain results that claim to be valid for all times. In the face of its object social science seems not at all to be able to reduce the complexity in the time dimension via the medium of validity. On the other hand, however, this reduction is necessary, if the existence of the action system science is not to be risked. We can now reformulate the above question: How can social science cope with the complexity in the time dimension? The answer is as simple as it is astonishing: by means of the concept of family resemblance. The result of its vagueness was the possibility to describe the contingency of social action and to "internalize" this contingency into the system of science. Therefore this vagueness is at the same time the open possibility to integrate future cases into the concept, i.e. to reduce the complexity in the time dimension.

For we can be sure that we will for instance call a game to be invented some time in the future a game, although we do not know anything about it now. The openness of the concept turns out to be its strength as far as the reduction of contingency in the time dimension is concerned. The contingency of open future is taken as openness into the concept itself, i.e. transferred from the time dimension into the object dimension. It becomes semantic contingency.

Our security henceforth to follow a rule in the application of the term game for instance is a security of pragmatics and not of semantics. It is based on the fact that we will change the meaning via the use, if this is necessary. Thus the security is based in the social community of the scientists. Thus the contingency is again transferred from the object dimension into the social dimension, where the scientists are responsible for it (cf. 2, § 76). Here another problem, which has not been mentioned yet, turns up. If we succeed in thus integrating the contingency into the concept at the "price" of a shift in the meaning of a word, then the changing relation of word and meaning reflects the interaction between object and concept. And this inter-

action is nothing else than the interaction between the researcher and his research object. The price we pay is the depiction of our history as a history of meaning, and that means of our form of life.

Such a depiction is possible within the clusteranalytic procedure of DAVTOLA, a procedure of non-disjunct classification (5, pp. 332-336). In it a new conception of science, open research, becomes evident paradigmatically.

6. The paradigm of open research

By open research I mean first a kind of research that imputes to its object as little theory as possible, and lets itself largely be governed by the object; it rather "snuggles up" to the object than forces it into a theoretical frame. After all I have said so far cluster analysis is such an adaptable procedure, in contrast to factor analysis, which also comprises different properties or objects and reduces given complexity, but at the same time presupposes resp. demands that the result is a linear shape. While cluster analysis visualizes an open field, factor analysis constructs a reduced abstract space. It does not transcend the space idea and is therefore theoretically finite.

Second, I demand that open research obtains open results. The results of science consist of conceptual systems and of laws. If these are to be open, it means openness for changing subject-matters, openness for slight deviations and exceptions and openness for variation. In the social sciences we always deal with exceptions, variation and evolution in the subject-matter. Let us look at an example. The empirically significant concept of democratic instruction is more than 50 years old. We still use it, and we use it relatively safely, although its full meaning is still unknown to us. Democratic instruction is an open concept of family resemblance. Therefore its meaning and its laws can only be investigated appropriately by means of cluster analysis. If democratic instruction were defined classically per differentiation specifics, then it is my thesis that either there would not be democratic instruction any more or the term would have disappeared from science. For it is certain that instruction has developed in the last 50 years in such a way that rigidly defined concepts (and laws) would not have been able to "survive". Interestingly enough it is just the results of research that this historical development can be traced back to (6). In action-relevant sciences we always find an interaction between research work and research object, wherefore we need flexible conceptual systems like those generated by cluster analyses.
MODELS FOR THE MEASUREMENT OF SIMILARITY

Outline:

0. Introduction

1. Psychological research of the perception of similarity

2. Measurement of similarity

3. The set-theoretical model of Tversky

4. Related models

The purpose of cluster analysis or automatic classification, respectively, is to find \( N \) clusters in a set of \( N \) objects \( O_1, \ldots, O_N \) based on \( p \)-dimensional vectors of values of variables \( V_1, \ldots, V_p \), which are given for each object. A cluster is defined as a subset of objects, which are more similar to each other than to any objects outside the cluster. From the definition of clusters it is self-evident that the choice of a suited similarity measure is of utmost importance in cluster analysis and that different similarity measures can lead to different classifications. It is generally assumed that similarity and distance are complementary concepts and that two objects with high similarity have a low distance and vice versa. By using transformations of the kind

\[ d = -\ln s \]

one can transform similarities into distances. In most cases metric distances are used in automatic classification, which are in accord with the axioms

1. Minimality: \( \delta(O_i,O_j) \geq \delta(O_i,O_k) \) for all \( O_i,O_j \)

2. Symmetry: \( \delta(O_i,O_j) = \delta(O_j,O_i) \) for all \( O_i,O_j \)

3. Triangle inequality: \( \delta(O_i,O_j) + \delta(O_j,O_k) \geq \delta(O_i,O_k) \) for all \( O_i,O_j,O_k \)

In (14) results from psychological experiments were reported, which show that the perceived similarity of human subjects need

Studies of how people rate similarity have been reported in a number of studies. For example, Tversky and Krieger (1971) found that people tend to rate two objects as more similar if they share a common attribute. This is consistent with the idea that similarity is based on common features. However, this may not always be the case, as people may also rate two objects as similar if they are both familiar or both unfamiliar. This is consistent with the idea that similarity is also based on familiarity or unfamiliarity.

References:


Joachim Krauth