The Notion of "Category": Its Implications in Subject Analysis and in the Construction and Evaluation of Indexing Languages

Mario Guido Barite

School of Library Science
University of the Republic of Uruguay

Mario Guido Barite is a professor and researcher at the School of Librarianship at the University of the Republic, in Uruguay, where he specializes in the Theory of Classification and Terminology. He is the author of six books and thirty articles about Librarianship and Information Science, and Law Documentation and Terminology.

Barite, M. G. (2000). **The notion of "category": its implications in subject analysis and in the construction and evaluation of indexing languages.** *Knowledge Organization, 27(1/2).* 4-10. 11 refs.

ABSTRACT: The notion of category, from Aristotle and Kant to the present time, has been used as a basic intellectual tool for the analysis of the existence and changeableness of things. Ranganathan was the first to extrapolate the concept into the Theory of Classification, placing it as an essential axis for the logical organization of knowledge and the construction of indexing languages. This paper proposes a conceptual and methodological reexamination of the notion of category from a functional and instrumental perspective, and tries to clarify the essential characters of categories in that context, and their present implications regarding the construction and evaluation of indexing languages.

Keywords: Categories, Classification theory, Indexing languages

1. Introduction

Recent literature concerning the topic of categories does not abound. Research in the Organization of Knowledge has firmly led its way in other directions (computing applications, machine indexing, terminology, thesaurus specialization vs. natural language, etc.). While categories appear to matter only to classificationists (i.e. to the limited group of those who elaborate and construct indexing languages and other conceptualization systems), sooner or later, most classifiers or indexers, have to assume the role of classificationists since the present state of indexing languages entails minor and major surgery be performed to adapt these languages to users' requirements.

The question proceeds: What do we designate when we make reference to categories? Eric De Grolier (1962) affirmed, some years ago, that there was no settled definition. Historically, the notion of category has had diverse conceptualizations, from the times of Plotinus and Aristotle to Kant and other modern philosophers (Hamelin and Husserl among others). It was Ranganathan who extrapolated the concept from Philosophy to the Classification of Knowledge and who elaborated a harmonic and reflexive conception of categories. Furthermore, to prove that categories necessarily are the foundation, not always visible, of any organizational system of knowledge, he constructed a system of classification, the Colon Classification, from his theoretical postulates.

In search of support to understand the concept, we refer to a good encyclopedic dictionary (Labor, 1990), which defines categories as "any of the more general characteristics or temperaments applied as a whole to all beings and to the other manners or modalities displayed by them". And it adds, under the entry SUPREME CATEGORIES: "Phil. A topic embraced by Metaphysics; it intends to systematically establish

more general modalities which globally include not only the entirety of beings, but also all their characteristics or properties, but they differ from essential categories in that each one belongs to one sector of reality".

As it may be appreciated, comprehension of this concept is neither simple nor easily accessible. The fact that in our literature, the notions of *category*, *characteristic* or *class* are sometimes used indistinctly even though they are so different, does not help to their characterization either.

We are also aware of some sort of provisional approach to the concept of *category* in the Theory of Classification. This approach is taken as conclusive and concluded by all of us – something like an unclosed reflection of Ranganathan's idea – which everybody, even Ranganathan himself, has assumed to be the definitive one. But, do we know for sure the applicability of such an abstract notion made in Organization of Knowledge? Have we managed to characterize, with extreme accuracy, the statements of such a conceptualization? Have we managed to establish, plainly and clearly, why and what for categories are useful in our disciplinary field?

There is a short story by Poe (1845), where Minister D. purloins an extremely important document from the Royal apartments. The police know this document may only be hidden at the Minister's house, and despite the fact that all of the furniture, joinings, rungs of every chair, boards, bedposts and paving bricks of the residence had been thoroughly scrutinized and even dismantled, the letter could not be found. The opinion of Auguste Dupin, a sort of scientific detective or scientist made into a detective, was requested and he found the document in the most unmistakable place: a modest card-rack of pasteboard situated where everybody could see it, expressing perhaps that "the mystery was a little too plain – or a little too self-evident."

The answers to our queries are also solved in the most evident way, from an analysis of the implicit discourse of Ranganathan and other experts, who apparently could not take the following step (the explanation of principles, statements and inferences regarding categories) because, maybe, they had it too near. The preceding has been said, it should be noted, with the most intellectually humble attitude we are able to hold.

2. Definitions

We should first make it clear that it is not possible to characterize categories in the Theory of Classification, taking on loan the definitions provided by Philosophy, Ontology or Metaphysics. They are mere and simple conceptual departure points to organize our specific scientific discourse. A first useful approximation tells us that categories are extremely general abstract expressions, so that they could be perceived in any entity, element or object. There is an implicit characteristic as well, following Aristotelian thought, which refers to the instrumental character of categories. Effectively, categories are used as tools to discover certain regularities of the material world, thus, all objects – at least all those belonging to the material world – have certain properties. As a result, *properties* is a possible category to analyze the material world.

Their abstract nature having been instituted, it should be added that categories are, in their basic nature, extremely simple notions – the most simplified ones we may obtain to analyze any phenomenon and place it in certain position with respect to an object.

In like manner, as categories express regularities, by force they also constitute conceptual structures with certain permanence and stability (though the result of their application to different objects may be variable), capable of having expressions in any object, entity or real or ideal being.

As a central core of our conception, we define categories as simplified abstractions that, with the strength of intellectual instruments, are used by classificationists to investigate regularities of objects of the physical and ideal world and for representing notions. This analysis and representation is performed with a view to logically organizing systems of concepts, sufficient for the organization of knowledge in general terms, and subject analysis or document classification in specific terms.

Within the Theory of Classification then, categories are only relevant as instruments of analysis and organization of objects, phenomena and knowledge. Although they keep their ontological essence, in our discipline categories are of interest not as elements of metaphysical speculation, but as levels or dimensions of analysis applied to the interior structuring of human knowledge and their most representative abstractions: concepts.

3. Usefulness

For the classificationist, the use of categories becomes a first line instrument in three precise activities mentioned in this paper:

i) design, planning and structuring of indexing languages or systems of knowledge (systems of classification, thesauri, taxonomies, etc.);
ii) modification or specification of classification tables;
iii) the evaluation and analysis of indexing languages and systems of concepts through a set of parameters capable of establishing the grade of reciprocal tension among related concepts and their relevance and validity.

For classifiers/indexers, the notion of category facilitates the subject analysis process and its symbolization, since it helps to establish correct precedence among several subjects in a document, as well as an adequate hierarchical arrangement to compose classification symbols or to correctly place all subject-matter in a chain indexing process.

Those responsible for classification systems usually suggest a standard order to name subjects, based on categories. Thus, in the Guide to use of UDC (McIlwaine, 1994), the recommendation is to apply the series "Thing – class – part – material – property – process – operation – agent – space and time". Therefore, if a document referring to "Material components of planes and flight operations" is to be classified, the series will be: "Planes – material components – flight operations".

4. Category, object and analyst

Strictly speaking, it is not possible to isolate the notion of *category* from those of *object* and *analyst*.

The analyst is the one who sets the basis for the analysis, and performs a subjective levelization, to have access to a specific reality of the object. He or she is conscious of the intellectual impossibility of embracing the knowledge of an object all at once and selects a distinct aspect of the object. In our discipline, classifiers and classificationists must consider all of the ways an author has approached a document, as well as their own subjective levelizations of analysis.

On the other hand, the existence of categories without an object of study is not possible, since applying categories in emptiness is senseless – the use of categories demands a frame of reference. The object or referent may be any object or entity, being or phenomenon, that admits analysis as something autonomous. All that exists or happens in the Universe qualifies for study, and the same occurs with the knowledge expressed in documents. The object has, therefore, a vast case study, as many ideal or material things occur in the world. Referents may be the city of Montevideo, the process of photosynthesis, Euclidean geometry, the human digestive tract, unicorns or the French Revolution.

Every object gives origin to infinite aspects for its analysis, given its intrinsic complexity. The French Revolution, for example, may be studied as a simple sequence of chronologically distributed facts, or in terms of its influence upon its historic times, upon the neighboring countries, upon the population of France, or upon the Arts and culture in general. The ideology of the French Revolution may be analyzed, or the gap existing between its postulates and its accomplishments; its causes may be analyzed, its links and consequences; or a sociologic, ideological or statistical study of its process may be performed.

The huge complexity of any object impedes its global, integral and complete analysis. There exist, in addition, object attributes that condition its study:

a) *Any object is naturally dynamic and mutable.* For that reason, for its analysis to be accomplished, the object must be captured at a certain time and abstraction from its reality is required at a given moment.
b) *The object may be real or ideal.* It may have existed, as may be corroborated by its existence registers (Einstein or the city of Pompeii), or maybe it only has an immaterial existence, not physical, due to its nature (theorems, the human unconscious, the Greek Gods). These particular characteristics seem to obstruct the analysis, since analysts are condemned to act by approximation. However, once conventions have been clearly established by consensus, abstract objects are easily systematized – after agreement has been reached regarding what a theorem is or certain chronological and factual conventions of the French Revolution – the difficulty of giving intellectual access to the concept diminishes.
c) *Some objects have delimitation problems.* Attempts to produce a definition usually create discrepancies and shades of meaning among experts, so much so that they may cause a certain aspect of the object to be placed within one category or the other. But we also have the difficulties posed by the concepts that do not attain con-

ventional agreement. To exemplify, think of the difficulty of approving by consensus the basic statements towards the definition of the concept *labor flexibilization* from the viewpoint of a sociologist with a Marxist orientation and another one of ultra-liberal ideas.

By association, terminologists know well that as many definitions of an object may arise, as there are theoretical, practical, semantic or ideological needs are intended to be met.

d) *A large part of the objects belong to, or occur in a phase of the time-space continuum, or rather flow along a section of that continuum.* Due to their mutating and dynamic nature, some objects achieve various configurations and undergo a double influence: that of the processes occurring as a result of the action of internal agents, and that of the processes caused by external agents. This double influence is the determinant of each specific configuration, since any object is, in a given time and in a given spatial situation, the synthesis of the impacts brought about by such agents.

As a matter of fact, many of the preceding theorizations preoccupy researchers more at an earlier stage of the scientific method than they do analysts of knowledge. Nevertheless, these theories provide us with an expressive picture of the inherent complexity of any analysis, and we should know them, to use with adequate methods to address the transfer of disciplines and their concepts from classification of knowledge to indexing languages.

5. Characters of categories.

In this section we believe we make a contribution towards the consolidation of our theoretical corpus through the transfer, to an explicit level, of different notions which were merely suggested, implied and even contradicted, in Ranganathian and post-Ranganathian thought. Unusual as it may seem since they are positioned at the base of Ranganathian theory, these notions have been neglected to some extent. Since it is established that in Science, even what is implied must be explained and justified, for this particular case we shall try to decompose the notion of *category* through inductive-deductive processes, in order to extract its most typical characteristics:

a) *Every category is a sectorial one.* Among the authors of our discipline, we have not found one who makes an expressed reference to the fragmentary or partial scope of the knowledge of reality offered by each separate category. Every category makes a division in a globality. As a result, the set of categories selected by an analyst should provide a complete representation of the object. But to achieve that would require that the set of selected categories encompass the total sum of possible levels of analysis. Ranganathan developed a collection of five categories to that effect: MATTER refers in its conception to the aspects of static analysis of an object; ENERGY corresponds to the object's dynamic study; TIME and SPACE refer to the object's position along those two coordinates; lastly, PERSONALITY may include both kinds of things and parts of things and operates as a residual category that absorbs those elements not contemplated by the other categories.

b) *Every category implies a specific level of analysis.* Because it has instrumental character, every category is functional and given its absolute degree of abstraction, it lies outside the real world. By reason of its functional character, the selection of a category always pursues an analytical aim, but only from a given approach, point of view or level of study. Thus, the ENERGY category attempts to establish, identify and decompose the set of processes undergone by an object, and simply that. Regarding Medicine, this category will enable the determination of signs and symptoms, the course of diseases, *etc.*

c) *Categories are levels of analysis external to the object.* We believe that the assumption that categories are constituted by object elements or components has erroneously spread, its roots proably going back to Aristotle. Notwithstanding, in our line of thought, when we establish as a postulate that every category is a level of analysis applied over the object, we naturally come to the conclusion that such categories are autonomous and external elements with respect to the object being studied. Consequently, TIME and SPACE categories are formed with totally external components, all that can be said in this respect is that they study the object in a given time or sphere of action. The fact is that their own instrumental nature makes categories akin to

laboratory equipment used by any scientist in his or her specific area. The opposing assertion would imply, for example, mistaking the cells a geneticist has been investigating for the microscope used to perform his or her research.

d) *Categories are mutually excluding.* If we accept that a category provides fragmentary information of the reality of an object, it should also be pointed out that each category envelops that sector exclusively and excludingly, ruling out any possibility of intersection between the level of analysis of one category with the level of analysis of another category. Jones gives apparently solid arguments to relativize mutual exclusion among categories. He expresses: "If a tree is examined it is found that *to be* a tree, as distinct from timber, it is essential for its roots to be in soil and for it branches to have access to air and light. If soil or air or light is removed there can be no tree – only timber: 'tree-ness' is energy dependent and tied to matter. As the combinations of energy and matter are only found under certain conditions (e.g. not above certain altitudes) trees can only exist in certain places. Further, individual trees are time dependent. Thus, 'tree-ness' (the personality of trees) is related to matter, energy, space and time and none of these elements can be excluded from it. Moreover, as all these elements are required it would be futile to attempt to assign especial significance to any one of them, such as time or matter" (158). Nevertheless, his proposal is debatable as soon as we discover that Jones has been more committed to finding examples of interdependence and non-exclusion in the real world and in nature rather than in the world of notions. The fallacy lies in denying mutual exclusion within reality, when it should be accepted in the level of abstractions. The reciprocally excluding character of categories is an inescapable need for their application, and it is useful in all types of indexing and creation of schemes.

e) *Every category is highly generalizable.* In the revision of literature of our specialized area, we confirm that in the broad application of categories, there is agreement among authors such as Mills (1964) "... concepts of high generalization and large application..."; Langridge (1977) "... the most general classes of phenomena", Wersig & Neveling (1976) "... a class of very high generality" or "a general facet applying to a lot of subject fields", and Buonocore (1976). If we depart from Aristotelian philosophy, and even Kantian philosophy, *category* – by nature and definition – should be applied indiscriminately to any object. Nonetheless, when the concept is transferred to the Theory of Classification, it seems clear that categories are instruments of high generalization, though not absolute. Conceivably, categories of larger generalization may be SPACE and TIME, and for that reason facets recognized as intrinsic usually appear in indexing languages as auxiliary or additional tables. As anyone who has organized a table of concepts may ratify, TYPOLOGY, (or TYPES) and PROCESSES are also categories applicable in a large degree to all disciplines and areas of knowledge.

The fact that some categories are highly generalizable does not imply that permanently and under any circumstances they must be used to construct an indexing language. There are areas of knowledge where the application of certain categories is not of any use. Every discipline has its own conceptual structure which pre-shapes categories to be used for its internal organization. The Colon Classification is expressive in this sense, since the facet formula rarely forms the complete PMEST sequence and does not correspond class to class.

f) *Every category may admit, with reference to an object, variable levels of subdivision.* Each level of subdivision is technically known as *characteristic*, *facet*, or *attribute*, through which a concept or object is subdivided; a homogenous set of foci being obtained in that way. Thus, if to the concept SOUTH AMERICA, the characteristic "by countries" is applied, we obtain a series with the name of all the states situated in that continent. The literature does not give very precise definition to the bond which joins and distinguishes the concepts *characteristic* and *category*. They are terms frequently used indistinctly as synonyms. But they are not synonyms. Though it is true that the characteristic also suggests a certain level of analysis of an object, its space is always involved in a more comprehensive sphere, that of a category.

The creation of a facet, when a characteristic is consistently applied, may eventuate in the frame of one category and only one category. Significant are the doubts that Ranganathan had, to assign either the formulation PERSONALITY, or the formulation ENERGY to certain facets, resorting to the simultaneous designation of two categories, {E} and {2P}, and leaving the classifier free to determine the applicable category in the context of the document.

The first statement we put forward to clarify this problem is: "Given an object, and a characteristic having been obtained for its analysis, the latter, in the level of analysis used, corresponds to one category and not more than one category". The following is inferred from this statement, by inversion of its meaning: "Given a category, *n* number of characteristics (and therefore, possible facets) are comprised by it, which will belong exclusively and excludingly to this category in the level of analysis used".

Let us confirm these postulates in a given case, to prove that for one category, a variable number of exclusive characteristics may always be identified: If the object were "humanity", in generic terms, the TYPES OF OBJECT category would allow the application of, among others, the following characteristics: "by ethnicity", "by nationality", "by social class", or "by gender". In the specific case, the TYPES category remains invariable, while the four identified characteristics are mutually excluding, and each of them is placed within the TYPES category in this level of analysis.

g) *Agreement has not been reached regarding a limited collection of categories.* Aristotle had originally recognized ten (substance, quality, quantity, position, possession, action, relations and others). Ranganathan reduced them to five, known with the acronym PMEST: personality, matter, energy, space and time. His followers have devoted study to either decomposing his categories (Husain & Khan,1990) or to establishing different repertoires. Thus, the Classification Research Group (CRG) of London proposed an order based upon the following categories: types of final product, parts, materials, properties, processes, operations, agents, space, time and form of presentation. Eric de Grolier distinguished among three constant categories (time, space and action) and seven variable categories (substances, organs, analysis, synthesis, property, shape and organization). I. Dahlberg proposed the following sequence: entities, temperament, properties, processes, phenomena, relations, spatial dimension and temporal dimension. Ranganathan himself may have contributed to that variety of positions, since in part I of his Colon Classification (1960) he identified the main facets of each class with normalized terms (Substance, Problem, Process, Organs, Parts, Material, Action, Utility, Operation, etc.) That were taken by other authors to designate new categories, arguing that the PMEST formula was insufficient.

In reality, the number of categories a classificationist is able to establish for his or her work shall increase inversely to the degree of generality of application aimed at in each category. Thus, the election of ENERGY will project a field of use relative to all dynamic aspects of the object. But, if inversely, the search is to specify certain dynamic aspects in "minor" categories, this will lead ENERGY to decompose in PROCESSES, OPERATIONS and PROBLEMS. There is no better alternative, only a decision linked to the utility it conveys for the analysis of objects.

6. Conclusions

It is possible that the conceptual and methodological revision we have tried regarding the notion of category demands more specific applications in indexing languages, to corroborate its adequacy and relevance. We also understand that a contribution towards its study from a functional-instrumental perspective may help to reconsider our systems of ideas and procedures regarding the construction and evaluation of indexing languages and the customary classification of documents.

We propose greater attention to this topic, because it involves essential theoretical-practical aspects for the reasonable command of the theory of concepts by specialists, as well as by lecturers of the corresponding subjects of the syllabi in our discipline.

Bibliography

Buonocore, Domingo (1976). *Diccionario de Bibliotecología.* 2ª ed. aumentada. Buenos Aires: Marymar, 1976.

De Grolier, E. (1962). *Etude sur le catégories générales applicables aux classifications et codifications documentaires.* Paris: Unesco.

Diccionario enciclopédico LABOR. (1990). Barcelona: Labor.

Husain, S. & K. Khan. (1990). Conceptual changes in the fundamental categories in C C 7. IN *Herald of*

Library Science, 29 (1/2). p. 43-50. These two authors propose that expressions of MATTER category are of three types: material of matter, property of matter, and method of matter. This idea was applied in different disciplines, and it is one of the novel aspects of the 7th edition of Colon Classification.

Jones, Kevin (1973). The environment of Classification: the concept of mutual exclusivity. IN *Journal of the American. Society for Information. Science.*, 24(2). p. 157-163.

Langridge, D. (1977). *Classificaçao*. Rio de Janeiro: Interciencia. p.36.

McIlwaine, L. G. (1995). *Guide to use of UDC.* Madrid, AENOR..

Mills, J. (1973). *The Universal Decimal Classification.* New Brunswich..

Poe, Edgar Allan (1845). *The purloined letter.* See http://bau2.uibk.ac.at/sg/sg.html

Ranganathan, S. R.(1960). *Colon Classification.* 6th ed. Bombay. *See* pages of part 1, with examples of each class.

Wersig, B. & G. Neverling (1976). *Terminología de la documentación = Terminology of documentation.* Paris: Unesco. p. 123-124.