

Reports and Communications

FID/CLA/LI: Compatibility Experience of Indexing Languages in the Education Field

The Working Group on Indexing Languages of the FID Latin American Commission (FID/CLA/LI) was created in 1980. Its main fields of activity are concerned with the compatibility of documentary languages and the co-ordination of terminologies.

The "Inventory of non-traditional documentary languages in Spanish and/or Portuguese" was elaborated by the Working Group on Documentary Languages REUNIBER 78 and published in 1981. This inventory showed that there exist several controlled languages which have been elaborated or are being used by different institutions and which cover common or adjacent subject fields.

In addition to the duplication of efforts this situation evidences, it also creates a problem of incompatibility among the different languages used for information storage and retrieval, even within the same subject field, in those institutions which render information services and wish to conduct information searches in the different systems.

The program of activities of the FID/CLA/LI Working Group was started by the performance of a survey on methodologies for the compatibility of documentary languages, and was followed by a compatibility experience among thesauri covering the same subject field, the education area being chosen for this project.

The objectives of this first project were to:

- 1) conduct a survey of the use and elaboration in Latin America of indexing languages in the education field, and
- 2) apply a specific methodology for the compatibility of documentary languages used in the region.

The results of the project were presented at the XVIII FID/CLA General Assembly, held in Buenos Aires on 9–10 December 1982.

The diagnosis

With the purpose of gathering information, the Group drafted a questionnaire and distributed 150 copies of it among the countries of the region. The aim of this survey was to identify documentary languages used for information storage and retrieval in information centers and specialized libraries of Latin America whose main subject fields were related to the education area or to a specific topic in this area.

Of the 150 questionnaires distributed, 60 were completed and sent back, 52 of them being relevant. 50% of the respondents were using non-traditional indexing languages, mainly thesauri, while various institutions were using both these languages and classification systems such as UDC or Dewey.

To these percentages we may add a 16% of respondents who were using classification systems only but which showed themselves very interested in adopting a thesaurus. This 16% evidences a growing tendency in the education field towards the use of non-traditional indexing languages; a tendency already noticed at a more general level through the Inventory of documentary languages.

The report showing the results of the project includes a table where we can see which specific languages are used in each country. Through this table we can also notice the great variety of tools for information indexing and retrieval that are being used by educational information units of Latin America.

The most used thesauri are: UNESCO: BIE Thesaurus on Education, Tesauro Colombiano de la Educación (Colombian Thesaurus on Education) of the Instituto Colombiano para el Fomento de la Educación Superior (ICFES), Multilingual EUDISED Thesaurus of the Council of Europe, Standardized Terminological List: Education Section of the Superintendencia de Educación de Chile.

27% of the respondents advised they were using languages developed by themselves. This indicates that there exist many different languages covering the same subject field, thus bothering to a great extent the compatibility of languages among the systems using them. Some 75% stated they were willing to make their languages compatible with those of other organisms, the exchange of information being their purpose.

We may add that 45% of the respondents were using Table 37 of the UDC, either exclusively or as a complement to thesauri. These institutions expressed they were for the compatibility of UDC with a non-traditional language.

The survey also made it possible to collect other types of data, such as kinds of materials indexed by the institution, categories of users and archive systems. With regard to the latter, it can be asserted that several institutions use languages (such as thesauri) suitable to conduct postcoordinated searches and use archive systems which do not permit postcoordination (sequential files).

Another important characteristic noticed through the survey was that almost all respondents maintain exchange relationships with other similar institutions or participate in educational information networks at a national or regional level.

The compatibility experiment

The Group conducted a bibliographic research project in order to determine the progress made with regard to the compatibility of indexing languages in similar subject fields, its purpose being to carry out a practical experiment applying a certain methodology for the compatibility of a set of documentary languages in the education field.

After analyzing the different proposals taken from the bibliography, the Group selected a compatibility method described by H.H. Neville in his article "Feasibility study of a scheme for reconciling thesauri covering

a common subject" (Journal of Documentation, vol. 26, No. 4, 1970, p. 313-319).

Neville's method is based on the assumption that "it is concepts that are indexed, the keywords being merely convenient though sometimes arbitrary labels for concepts", and that "broadly speaking, thesauri covering the same subject must cater for the same concepts, although they may use quite different keywords to label them".

His method comprises two stages:

1) Given several thesauri, thesaurus A is taken as the "source thesaurus" and all its keywords are given code numbers and reconciled with all the other thesauri.

2) Thesaurus B is now taken as the source thesaurus, and all those keywords in it which were not dealt with in stage 1 are now reconciled in the same way with all the other thesauri, including A. The same procedure is applied to reconcile the remaining thesauri.

Neville shows eleven types of problems and proposes a method of reconciliation for each. The Group considered cases 1, 2, 3, 4 and 7 for the compatibility experiment, and employed the following thesauri: UNESCO: BIE, EUDISED, Colombia's, Chile's and Table 37 of UDC, which were the most used documentary languages in accordance with the diagnosis. The UNESCO: BIE thesaurus was taken as the source thesaurus and reconciled with the other ones.

The Group selected a sample of forty (40) concepts represented by descriptors of the UNESCO:BIE thesaurus. The said descriptors were subsequently numbered and reconciled with the other documentary languages already mentioned. Neville's methodology needed some adaptation when reconciling UNESCO:BIE thesaurus with the UDC. Besides, different versions of Table 37 and other tables of the UDC had to be consulted.

The report on the project describes the procedure applied to each term in the sample. The results of the compatibility experiment were organized in the form of an equivalence table which shows how each term is displayed in the five thesauri reconciled by the Group. Every term is accompanied by a compatibility code, all codes forming an intermediate lexic on through which any of the systems can consult the others by using its own indexing language.

The following aspects were considered by the Group for the compatibility experiment:

a) *subject field*: the education field covers several disciplines which fall within the social sciences area. Besides, polysemy problems are frequently observed when analyzing educational terminology; the same term may have different meanings in different countries or regions, while different terms are used to represent similar concepts. To this we may add that, although nearly all the region shares a common language, every country has its own educational system and generates very specific terminologies.

b) *participating languages*: the Group analyzed the problems posed by the different levels of precoordination of the languages, the scarcity of their entry vocabularies (non-used terms which refer to descriptors and

UDC alphabetical index), the inadequacy of their scope notes or lack of accuracy, and the difficulties encountered when reconciling UDC with the other languages.

c) *methodology for reconciliation*: the Group also examined the percentages of occurrence in each of Neville's cases and also the percentages of non-correspondence.

Conclusions

The analysis of the situation in Latin America shows that there exists a great variety of documentary languages which are being used or elaborated by the different institutions in each country.

It would be necessary to reconcile existing documentary languages if the different institutions wish to carry out the storage and retrieval of information in a co-operative way.

The Group considered the methodology for reconciliation to be adequate and feasible for the compatibility experiment, which does, however, require considerable human effort.

This effort is justified if a co-operative state is achieved among various institutions for the storage and retrieval of information, especially if automatic media are to be employed.

The special characteristics of educational terminology must be taken into account. Besides, each educational system presents its own characteristics and poses additional problems and differences to those observed in terminologies on social sciences.

On the basis of the results of the diagnosis and the compatibility experience, it is possible to make the following suggestions and recommendations about the compatibility of existing documentary languages in the education field in Latin America:

1) Thesauri and languages of a general type must be analyzed in the first place, incorporating more specific languages afterwards. Among the latter we may find microthesauri and thesauri having a core in common with the general system, or other subjects which perhaps must not be reconciled (for example, Tesauro de Formación Profesional CINTERFOR - thesaurus on professional training -).

2) The task may be initiated by a group of institutions by selecting a common reference language and then ascertaining equivalences and differences with their own languages. The said reference language might be the UNESCO: BIE thesaurus since it is widely known and presents a terminology which does not contain regionalisms, etc. as it has been elaborated by an international organism.

3) These institutions must provide specialists in the education and information fields who must also be familiar with the scope and use of the terms at the institutions they belong to and at a national level as well.

4) A first stage would comprise the determination of equivalences of their languages with the reference one, thus producing a table of reconciled terminologies.

5) During a second stage the group in charge of the project would use the compatibility codes and the solutions proposed in Neville's methodology. This would

produce a reconciled system capable of being applied to an automatic and co-operative system for information storage and retrieval which any participating institution can refer to at any time to provide information or ask the system for information through its own languages. Compatibility codes would make it possible to automatically determine equivalent terms.

This activity is fully described in the following article: *Grupo de Trabajo FID/CLA/LI – Informe de Proyecto: lenguajes documentarios en el área educativa en América Latina: diagnóstico y experiencia de compatibilización* (Report of the project on: documentary languages in the education field in Latin America: diagnosis and compatibility experience). Ana M. Sanllorenti and Mercedes Patalano. Buenos Aires, CAICYT, 1982. 77 p.

Project for the compatibility of educational terminology in Latin America

On the basis of the experience obtained through the project previously described, a new project has been designed and is being carried out with the following characteristics:

Objective: determination of equivalent terms among the terminologies used by different Latin American countries in the education field.

Description: the project aims to facilitate the development of existing indexing tools and their compatibility through the elaboration of lists of equivalent terms indicating similar concepts on education in different Latin American countries.

Stages

- 1) On the basis of the diagnosis on documentary languages in the education field in Latin America, identification of institutions (information units and research institutes) and individuals interested in collaborating with the project.
- 2) Identification of bibliographic resources (documentary languages, glossaries, official documents containing educational terminology) which will constitute the working material of each country for the project.
- 3) Selection of a reference language and determination of terminological differences in each country. These differences will then be sent to the Coordinating Center.
- 4) On the basis of the information provided by each country, elaboration of a document which will contain tables of equivalent terms. This task will be handled by the Coordinating Center.
- 5) Distribution of the document among the participants for comments and proof-reading.

Ana Maria Sanllorenti

Ms. A.M. Sanllorenti, Coordinadora FID/CLA/LI
Consejo Nacional de Investigaciones Cient. y Tecn.
CAICYT. Moreno 431/33, 1091 Buenos Aires, Argentina

Classificatie Stichting

The Netherlands Classification Foundation called "Nederlandse Stichting voor Classificatie en andere Ontsluitingsmethoden" (NCS) is sponsored by four organisations in the field of documentation and information, namely

- (1) the Dutch member of the FID, the Nederlands Orgaan voor de Bevordering van de Informatieverzorging (NOBIN)
- (2) the Dutch Association of Librarians, Documentalists and Literature Investigators (NVB)
- (3) the Dutch Association of Company Archivists (NVBA) and
- (4) the Association for Documentation and Administrative Organisation (SOD).

The Council of the Foundation consists of seven people, each of the organizations named is represented. As of Jan.1, 1984 the following persons are members of the Council: J.J.Kroese (NOBIN), chairman; A.Hartkamp (SOD), secretary; W.Kok (NVBA), Dr.H.A.van Leeuwen, Drs.P.A.Tichelaar, A.Vasiljev, and G.J.Wijndands (NVB). Any correspondence should be directed to: Secretariaat NCS-Bestuur, p/a A.Hartkamp, Gemeente Hellendoorn, Willem Alexanderstraat 7, NL-7442 MA Nijverdal. (Source: Open 16 (1984) No.2, p.107).

EURALEX founded

At the close of the LEXeter 83 Conference, the European Association for Lexicography was founded for the purpose of promoting "scholarly and professional activities related to dictionary-making, including the training of lexicographers".

This was the result of developments in a number of fields concerned with the production and use of dictionaries. Several different aspects of lexicography have recently received the attention of applied linguists, computer specialists, terminologists and others, but none has had the benefit of the unified perspective which the new association hopes to offer. EURALEX plans a range of publications and meetings in furtherance of its declared aims. Contacts and cooperation with other bodies (such as the Dictionary Society of North America) will be sought, and fresh channels of communication will be explored.

An International Executive Board has started work. Composed of Gabriele Stein (Hamburg) as Chairman, Noel Osselton (Leiden) as Vice-Chairman, Reinhard Hartmann (Exeter) as secretary, Frank Knowles (Aston/Birmingham) as Treasurer and five other members (S.Atkins, T.Cowie, R.Ilson, T.Magay, A.Zampolli).

There are four categories of membership: Active Membership for those engaged in lexicographical work, Student Membership for those in training, Associate Membership for other interested individuals or institutions, and Honorary Membership in recognition of outstanding achievement. Those who would like to apply for membership or receive a copy of the EURALEX Draft Constitution may do so by writing to: Dr. R.R.K. Hartmann, University of Exeter, Devon EX4 4QH, England.