A Universal Classification System Going Through Changes

Victoria Frâncu
Central University Library of Bucharest, Romania

Abstract

In the early 1990s, subject indexing with classification codes from the Universal Decimal Classification in academic libraries of Europe proved insufficient. With the advent of OPACs for online searching of literature, the possibilities of searching for subjects in natural language, offered by a majority of OPACs, looked much more attractive to the users and as well to the indexers. Thus, a rapid change to indexing with UDC and keywords instead of UDC numbers alone was decided. Currentness, precision and more importantly, user-friendliness were strong advantages offered by keyword indexing and searching. But the larger the dictionary of keywords, the more problematic are the consequences on information scattering given lack of control on terms. The present paper describes the advantages of the UDC in indexing by presenting some of the devices it is provided with: subdivision by analogy, common and special auxiliaries, use of synthesis, use of connecting symbols. The solution of indexing with both UDC notations and words from a thesaurus based on UDC was prompted by some other characteristics of the schedules: semi-faceted classification system, hierarchically organized, rich in terminology, providing consistency and control of notations. The methodology used in building the thesaurus follows the international standards (ISO 2788 and 5964) where some principles have been added, relative to the specific aim of harmonizing a classification structure with that of a thesaurus. Compatibility and translatability issues are also considered, and some problems arising from them are treated in detail. The problems discussed are illustrated with examples.

Introduction

The idea of going from free indexing to a controlled vocabulary based on the logical structure of UDC proved to be the solution to the problem of enhancing subject searching in online catalogues. In the beginning of the 1990s, a number of research projects on this issue were carried out (eg. Riesthuis and Bliedung, 1990; McIlwaine and Williamson, 1995; Riesthuis, 1995). These studies strongly recommended the solution of indexing with descriptors derived from UDC. These studies were preceded by much earlier ones dating from the 1960s (Ungarian, 1966) and the 1970s (D'Haenens and Lorphèvre, 1974), investigating the conversion of the alphabetical index of UDC into an alphabetical thesaurus. In the recent years, the subject has

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

gained particular interest along with the re-evaluation of the qualities of UDC as a unique organizer of human knowledge (Frâncu, 1996; Williamson, 1996; Scibor, 1997).

The project discussed in this paper was previously presented in the 4th ISKO Conference in Washington, DC, in 1996; by that time it had only been started and tested on the structure of Class 8 for linguistics and literature of UDC. The initiative gained supporters and developed into the ambitious plan of turning the whole structure of UDC into a thesaurus. In the meantime, other subdivisions of the schedules have been added to Class 8 (02 - Librarianship, 1 - Philosophy, 2/245 - Religion, 57/59 - Biology), providing evidence of the qualities of UDC as a basis of thesaurus structure and terminology. Some shortcomings have had to be surmounted as they were, but the end products are printed and used by now enabling us to draw conclusions on the feasibility of such a project.

UDC – a sophisticated indexing and retrieval tool

Historical background

Originally intended to systematically organize a comprehensive listing of everything that had been printed at the turn of the last century, the Universal Decimal Classification (UDC) came to be and still is the world's primary multilingual classification scheme for all fields of knowledge. That listing was by its Belgian authors - Paul Otlet and Henri Lafontaine – titled as *"Répertoire bibliographique universel"*. The system by which they organized the entries was an adaptation of Melvil Dewey's decimal classification scheme. They expanded its coverage and increased its flexibility by adding auxiliary tables and some synthetic devices (McIlwaine, 1993).

The first edition was published between 1905-1907 under the title *"Manuel du répertoire bibliographique universel"* and consisted of about 33,000 notations. The maintenance and revision of the scheme became permanent activities and they were taken over by Institut International de Bibliographie later known as Fédération Internationale de Documentation (FID). In 1991 FID transferred the ownership rights and responsibilities over the UDC to the Universal Decimal Classification Consortium.

The rapid growth in the volume of human knowledge imposed the necessity of revising, changing and sometimes restructuring whole classes of the UDC scheme. Formerly shaped as a periodical publication - issued irregularly and rather seldom -, *"Extensions and Corrections to the UDC"* (E&C) is the result of remarks and suggestions made by the users of the classification scheme and directed to the body entitled to maintain and develop the system. From the early 1990s, Extensions and Corrections became an annual publication. Together with the revised UDC tables, it

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

additionally contains articles of professionals as much as lists of recent publications concerning UDC and UDC-related topics. Beginning with the 14th volume, major revisions have been done for classes like: 8 - Linguistics. Literature, 9 - History, 53 - Physics, 544 - Physical Chemistry, 2 - Religion (proposed revision), 61 - Medical Science (proposed revision), 004 - Computer Science as much as to the Common Auxiliary Tables of Language and Place.

Soon after its foundation, the UDC Consortium decided to create and develop a database called *Master Reference File* (MRF) containing the UDC scheme in machine-readable form. This was designed to serve as a basis for the further revisions and extensions and as a primary source for printed editions and other services. MRF was created in 1993 and it consists of about 62,000 numbers with their complete descriptions, annotations and examples of use in English. The main source of MRF is the International Medium Edition in English published by the British Standard Institution in 1985.

Structure and characteristic features of the UDC

UDC is a general classification scheme like Dewey Decimal Classification (DDC), Library of Congress Classification (LCC) or Bliss Bibliographic Classification (BC) (McIlwaine, 1993: 8-10). It is general because all fields of knowledge are included in it and it may be applied in collections covering the whole of knowledge.

There are two major characteristics of UDC which potentially allow for mapping it to a thesaurus structure:
(i) UDC is an *aspect classification* meaning that phenomena are subordinated to the aspect from which they are taken, thus:

'Rabbit' will be found in nine occurrences in MRF, under Palaeontology - 569.32, Zoology - 599.325, Agriculture - 631.225, Animal husbandry - 636.92, Hunting - 639.112, Leather industry - 675.318 and Textile industry - 677.354.

The problem of ambiguity of the concepts in the tables are automatically solved via class marks. Using thesauri derived from such a classification system, when homonyms occur, disambiguation can be achieved by combination of descriptors in order to show the aspect or field to which the concepts belong, thus:

Symbolism in literature
  USE: Symbolism + Literary trends
Symbolism in art
  USE: Symbolism + Art
Religious symbolism
  USE: Symbolism + Religion

This is different from the case of DDC where the notational synthesis provide hierarchical links to different aspects together with the facet indicators (Mitchell,

ISSN: 2324-9773

1997). The standard notation for 'Bears' 599.78 can provide links to broader or narrower concepts, for instance:

Bears	
animal husbandry	59**9.78**
big game hunting	636.**978**
technology	639.97**978**
resource economics	333.95**978**

If we take the 'Rabbit' example from UDC and compare it with the 'Bears' example from DDC, we shall see there is no relation between different aspects of the entity we consider in the former range of classification notations. The numbers are unique and there is no sequence of digits saying that we speak about the same entity as in the latter case is. Yet, references are provided to class numbers where the same concept is represented though in different aspects or disciplines.

(ii) UDC is a *hierarchical classification* subdivided into logical components by the successive application of the principles of division. These may be either *generic* or *whole/part* (McIlwaine, 1993).

The *generic relationships* link the class with its members or species (mostly in biology but throughout the whole classification as well).

Example:

Adjectives	81'367.623
Concrete adjectives	81'367.623.1
Abstract adjectives	81'367.623.2
Denominative adjectives	81'367.623.3
Numerative adjectives	81'367.623.4
Possessive adjectives	81'367.623.6
Demonstrative adjectives	81'367.623.7
Interrogative adjectives	81'367.623.8
Indefinite adjectives	81'367.623.91
Relative adjectives	81'367.623.92

The last two subdivisions of the above example have more digits because of the decimal principle on which the UDC is based: two more entities being necessary and one more digit left, the solution was to further subdivide the last existing one, i.e. 81'367.623.91 and 81'367.623.92 for Indefinite adjective and Relative adjective respectively. This applies also in the Literature section of Class 8 with the special auxiliaries for Literary genres.

ISSN: 2324-9773

The *whole/part* or *'part of'* type of relationship may apply to parts of the body or geographic location. This can result in a number of subordinated classes called chains.

e.g. Europe - South-Eastern Europe - Romania - Transilvania - Sighisoara

They can also generate a number of coordinated classes called *arrays* (sets of mutually exclusive classes derived from the application of one specific characteristic of division)

e.g. English poetry - English drama - English fiction

Another advantage of the hierarchical structure of UDC is that it can be successfully used in information searching and retrieval. A particular notation can retrieve several subdivisions when truncated. Buxton (McIlwaine, 1993) gives as example, '656.2$' which will retrieve among others:

656.2	Rail transport. Rail traffic
656.21	Operation of railway tracks, buildings, stations
656.22	Commercial organisation of railways. Train services
656.222	Running of trains
656.223	Use and distribution of rolling stock
656.224	Passenger train services. Including: Departure. Arrival. Control
656.25	Safety measures. Signals
656.27	Ancillary operations. Services

In terms of structure, UDC consists of two types of tables: main tables or schedules and auxiliary tables.

The *main tables* contain the multitude of disciplines of knowledge grouped in 10 classes (from 0 to 9) hierarchically subdivided. The classes in UDC are similar to those in DDC with one exception: the discipline in Class 4 of Dewey's classification - Linguistics - is grouped together with Literature in Class 8 of UDC, Class 4 being vacant in the latter.

The *auxiliary tables* indicate recurrent characteristics (facets) of the subjects denoted by the main number such as: language or physical form of the document, historical period, geographical location, nationality or ethnic aspects. The auxiliaries fall into two categories: *common auxiliaries*, applicable throughout the main tables and *special auxiliaries* restricted to particular classes.

The following examples show some of the most used common auxiliaries, the symbols they are recognized by and a few examples [1] to illustrate their use:

- =... *The common auxiliaries of language* denote the language of the document whose subject is represented by the main number:
  e.g.

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

030=**111**	→ Encyclopaedia in English
811.135.1'243(075.8)=**14**	→ Handbook of Romanian as a second languagfor students, in Greek language

- **(0...)** *The common auxiliaries of form* denote the bibliographic or physical form of the document. They should not be used to denote the subject.
  e.g.

53**(038)**=135.1	→ Dictionary of physics
75(492)**(084)**	→ Dutch painting album

- **(1/9)** The common auxiliaries of place denote the geographic location or aspects of space the main number refers to.
  e.g.

39**(498)**	→ Romanian ethnography and folklore
774**(73)**	→Cinematography in the United States

- **(=...)** The common auxiliaries of ethnic grouping and nationality denote the ethnic aspects of the subject represented by the main number.
  e.g.

81**(=411.16)**(498)(092)	→Biographies of Jewish linguists from Romania

- **"..."** The common auxiliaries of time denote the time or period of time to which the subject denoted by the main number refers.
  e.g.

94(498)**"1848"**	→The year 1848 in the history of Romania
355.48(4)**"1939/1945"**	→Military campaigns in Europe in the Second World War
75(492/493)**"16"**	→Flemish painting in the 17th century

- **-05...** The common auxiliaries of persons are used to emphasize aspects concerning the persons involved in the subject denoted by the main number with respect to age, sex, social group, class or profession .
  e.g.

616**-053.2**	→Pediatrics
821.135.1-1**-055.2**(082)	→Anthology of Romanian women poetry

The *special auxiliaries*, unlike the common ones, apply only to some of the classes. They can be preceded by **.0** (. nought), - (hyphen) or ' (apostrophe). There are special instructions in the UDC tables provided for their use within each class.

e.g.

616.7**-006-07**	→Diagnosis in the neoplasm of the bones
811.111**.02**	→Literary trends in English literature
811.134.2'**282**	→Dialects of Spanish

ISSN: 2324-9773

According to the succession of numbers *coordination* is expressed in different ways in UDC. Denoting the relation between subjects or aspects of subjects it is graphically shown by signs such as: “+” (plus), “/” (stroke), “:” (colon), “::” (double colon).

- “+” *Addition* - links two or more separate numbers
  e.g.

(44+460)	→France and Spain
(47+57)	→Former Soviet Union
662+669	→Mining and metallurgy

- “/” *Extension* - links two or more successive numbers to indicate a broader subject or a range of concepts.
  e.g.

(7/8)	→North, Central and South America
592/599	→Systematic zoology
23/28	→Christianity. Christian Religion

- “:” *Simple relation* - is used to relate two or more subjects but more restrictively than the addition.
  e.g.

17:77	→Ethics in photography
327(73:47)	→International relations between USA and Russia

- “::” *Double relation* or *irreversible relation* - is used instead of colon to indicate that the concept following it is in a subordinate relation with that preceding it. It is used to fix the citation order in a complex notation (McIlwaine, 1993: 38).
  e.g.

575::576.3	→Cytogenetics
77.044::355	→War photography

Indexing with UDC

It has often been said that UDC can be used with great relevance in indexing, but using it for information retrieval is not easy; it is a tool for indexers or expert users. Despite its numerous advantages, UDC is generally considered less effective as an information retrieval tool than the information languages based on the principle of coordinate indexing, i.e. keyword and descriptor languages, especially in online information retrieval (Scibor, 1997). We shall see later that there is a solution to make UDC more user-friendly; but let us first consider some examples of the way UDC can be used in indexing.

ISSN: 2324-9773

Les relationes hôtes-parasites dans le modèle Téléostéens-Métacercaires de Labratrema minimus (Trematoda bucephalide) / présenté par Elisabeth Faliex . - Grenoble: Atelier National de Reproduction des Theses, 1991

UDC notations:	*Descriptors:*
578.23:[597.5:576.8**95.122**(043)	Parasitology
578.23:[576.8**95.122**:597.5(043)	Teleostei (Fishes) Trematodes(Worms)

Example 1. Use of subdivisions from a different class according to instructions given in the tables

In this first example, not the easiest one to index, the concepts we have to denote by classification notations belong to Classes 57 and 59. They are: 'relations between virus and host cell' (578.23), 'animal parasitology' (576.89), and two species of animals - broadly speaking - 'teleostei' (597.5) and 'trematodes' (595.122). We need to connect these notations in such a way that the subject is coherently represented. But we see an indication in 'animal parasitology', saying that 576.892/.899 is subdivided like 592/599. Therefore, we have to go from Biology to Zoology and take over the right notation for the subject of parasite in our document.

Siebenburgisch-Sächsisches Wörterbuch : mit bentzung der sammlungen Johann Wolffs / Ausschuss des Vereins für Siebenbürgische Landeskunde. - Berlin: Walter de Gruyter

UDC notations:	*Descriptors:*
811.**112.2'28(498.4)(038)**	Dialectology German language Sachs Romania Transilvania Dictionary

Example 2. Use of synthesis, common auxiliaries of language, place and form

This example illustrates the use of synthesis in Class 8; the special auxiliary '28 can be attached to any of the individual languages needed or to 81 - Linguistics, if we deal with a linguistic study on dialectology. Further, the example shows the use of the common auxiliary of language, which is =112.2, German in this case, the common auxiliary of place (498.4) denoting the Romanian historical province of Transilvania and the common auxiliary of form (038) for dictionaries.

Studies in Pre- and Protomorphology . - Wien: OAW, 1997. - ISBN 3-7001-2654-9

UDC notations:	*Descriptors:*
81'366'276.3-053.2	Sociolinguistics Morphology Usage of language
372.46-053.2	Children Rudiments of speech
159.922.7:372.46	Child psychology

Example 3. Use of different common and special auxiliaries in one complex classification notation

ISSN: 2324-9773

The last example demonstrates the possibility of using more than one kind of auxiliary in one and the same notation. In the first UDC notation we have represented according to the captions of its component parts:

'366	Morphology
'276.3	Language or idiom of a particular age group or sex
'276.3-053.2	Infant or child talk

Finally, we may say that all the previously mentioned devices and additionally the remarkable richness in terminology of UDC, ranging from very broad concepts defining disciplines of knowledge down to narrow, very minute details and aspects of phenomena, can be advantageously re-evaluated.

Class 8 of UDC and its adequacy to thesaurus construction

Translation as a basic principle in information transfer

It has been frequently pointed out that indexing and searching involve each a process of translation (Iivonen, 1996; Maniez, 1997). Indeed, this is what the indexer is doing when turning the result of concept analysis over the subject of a document into the elements of the information language used in indexing. In the same way, the information query is translated by the searcher from his own discourse according to his information need into indexing terms (Fig. 1).

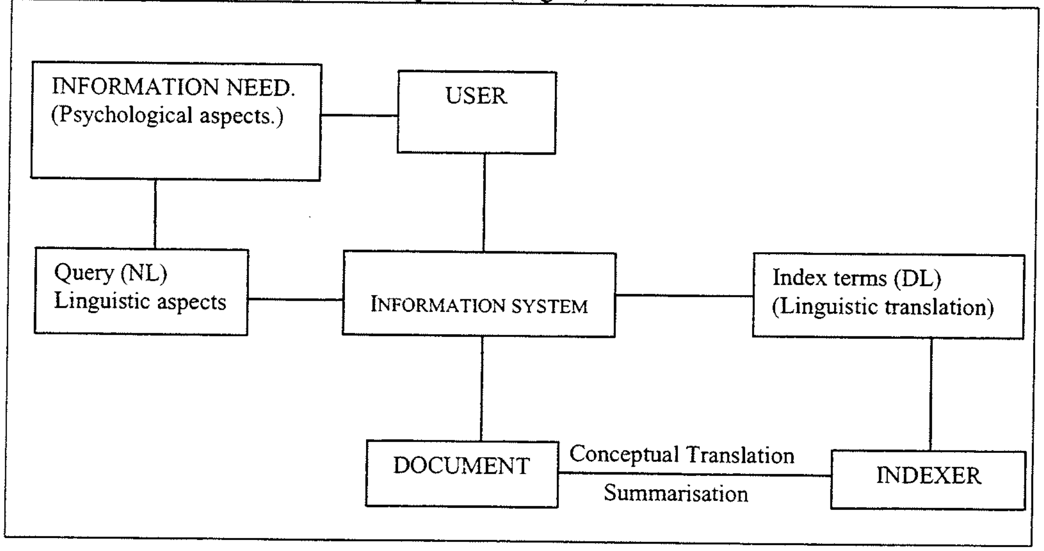

Figure 1. Translation processes involved in information transfer

The natural language (NL) terms in the discourse of the author are submitted by the indexer to a process of conceptual translation during the formulation of the 'Topic

ISSN: 2324-9773

Statement' (Hutchins, 1975). This is done by summarizing the main points and arguments, thereby unraveling the primary semantic thread of the text. Once formulated, the Topic Statement is turned into elements of the documentary language (DL) - either a classification system or a postcoordinate system - by linguistic translation. At the other end of the process of information transfer, the 'Topic Request' turns the information need of the searcher (user) from natural language terms in which he formulates his query into indexing terms belonging to a particular documentary language. This is inevitably less precisely formulated than the Topic Statement of the author and depends to a great deal on the level of knowledge or the searcher, on pragmatic influences as much as on his linguistic skills. The extent to which the indexing formulas and the search formulas correspond with each other have greatly impacts the predictability, and hence, the effectiveness of the information language.

Conversion and compatibility

The translation issues involved in information transfer presented above are of minor relevance to our approach unless we consider the level of specificity of the documentary language used compared with the category of users for whom this documentary language is intended.

What really matters is to what degree the classification system notation can be mapped into terms or formulas going to be used in postcoordinate searching? And then, what are the requirements needed by a universal classification system of such specificity as UDC to be successfully used in an online catalogue?

Glushkov et al. (1978) distinguish between *semantic compatibility* and *structural compatibility.*

The semantic compatibility takes into account the body of knowledge, or the discipline to which the information languages considered refer. The authors go further explaining that semantic compatibility can be reduced to lexical, paradigmatic and syntagmatic compatibility. In other words, the compatibility if a function of the representation of objects, the hierarchical relations and the non-hierarchical relations recognized. This implies that we have compatibility if whatever we express by a classification notation like 811.135.1'282'373.2 can be expressed without loss of meaning in words belonging to a natural language taken from a thesaurus: 'Romanian language', 'Dialects', 'Onomastics'.

The structural compatibility defined by the above mentioned authors, can be reduced to morphological compatibility (similarity in the structure of terms) and syntactic compatibility (similarity with respect to the structure of groups of terms or phrases). This point will be discussed later when translatability problems between the three languages will be presented.

ISSN: 2324-9773

According to Scibor (1997: 201), there are three methods of using UDC in information retrieval systems:

a) use of full (compound or complex UDC notations);
b) retrieval according to a subset (or subsets) of characters contained in a full UDC number;
c) use of verbal equivalents to UDC numbers.

As the first two methods are not of primary importance here, we shall consider the third. The solution adopted in order to make searching in our classified catalogue easier and more user-friendly was to connect each subject notation assigned to a document with one or more descriptors or descriptor formulas. In this way, a document indexed with 73(498) Brâncusi,C.(084) will have as descriptors: 'Sculpture', 'Album', 'Romania' and 'Brâncusi, Constantin'. This method has been suggested by several authors, together with more sophisticated approaches such as:

1) - verbal equivalents to UDC numbers (and also their synonyms and quasi-synonyms) serve as a mere interface and are automatically translated by the computer program into UDC numbers (Scibor, 1997:201);
2) - UDC numbers are used postcoordinately when given as subject representation; instead of 316(02.053.2) as a whole, the parts are given separately in the record; the request 3$ and (02.053.2) will retrieve the documents sought for;
3) - the full descriptions of the UDC numbers in the tables are used for each subject notation (Riesthuis, 1997:139).

In the first of these three methods, retrieval is transparently carried out by the computer, working from the UDC numbers applied in a search without the searcher being aware that the retrieval was not done according to his search expression. This means very complicated programming and massive work for software designers. The second and the third methods need a rather complicated treatment of factoring of the complex UDC notations by means of algorithms. These algorithms are specific to each type of notations and their connectors such as: splitting algorithms, algorithms for alphabetical additions, algorithms for parallel subdivisions, algorithms for notations with special auxiliaries.

Going back to our approach, the next step was to control the terms assigned to the records indexed with UDC numbers. The idea of the thesaurus based on UDC was therefore born.

Thesaurus structure

The thesaurus based on Class 8 of the UDC tables is built according to the rules of ISO 2788 (ISO, 1986), combined with those of ISO 5964 (ISO, 1985). From paradigmatic point of view, the thesaurus is organized in two parts: alphabetic and systematic. Apart from Class 8, the thesaurus contains a considerable part of the table for Common auxiliaries of language (more than 380 descriptors). The languages were

ISSN: 2324-9773

considered as an important component of both the disciplines of linguistics and of literature.

The *alphabetical display* of the thesaurus contains all the entry terms no matter if they are descriptors or non-descriptors. A descriptor group or a thesaurus entry consists of:

a) the entry term - descriptor - in all three languages with the first language term in upper case;
b) the 'scope note' (SN) if considered as necessary, providing either an explanation of the meaning of the term or instructions of use;
c) the UDC number the descriptor was derived from;
d) the 'used for' (UF) term providing a reference to the non-descriptor
e) BT, NT and RT.

Non-descriptors are given in lower case and they are provided with 'see' reference to the preferred term.

The *systematic display* is organized according to the standard filling order of the UDC schedules. Even though non-descriptors are not mentioned, they appear in the descriptor group introduced by UF thus making a link with the alphabetical part of the thesaurus (Fig. 2).

The thesaurus may have three variants with each of the three languages involved as a main entry language. What is really needed (and this will be corrected with the next edition of it) is an alphabetical index of the descriptors in English, French and Romanian as an intrinsic part of the thesaurus.

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

Alphabetical display	Systematic display
Colloquial language	
See: FAMILIAR LANGUAGE	
	81'286
GEOGRAPHICAL LINGUISTICS	GEOGRAPHICAL LINGUISTICS
F: Linguistique géographique	F: Linguistique géographique
R: Lingvistica geografica	R: Lingvistica geografica
SN : Used for the distribution of dialects in certain geographical regions	SN : Used for the distribution of dialects in certain geographical regions
UDC: 81'286	
UF : Dialectal geography	UF : Dialectal geography
BT : Dialectology	BT : Dialectology
	81'342.1
TIMBRE	TIMBRE
F: Timbre	F: Timbre
R: Timbru	R: Timbru
UDC: 81'342.1	
SN : Used together with another descriptor to denote the field (e.g. Timbre + Acoustic phonetics)	SN : Used together with another descriptor to denote the field (e.g. Timbre + Acoustic phonetics)
BT : Acoustic phonetics	BT : Acoustic phonetics

Figure 2. Alphabetical and systematic display of the thesaurus

There is a critical problem to be discussed here: should a universal thesaurus be one coherent universal thesaurus or a conglomerate of autonomous thesauri built along the same guidelines? If the first approach is considered, then very well defined tasks among the thesaurus designers have to be agreed upon, and tremendous intellectual efforts for working together simultaneously are needed. Problems of integrating and harmonizing concepts and disciplines will arise and they will have to be solved the moment they occur.

The second approach has different implications: it requires harmonizing work after the component microthesauri are made. The problem of polyhierarchies will then come up. Our solution was to introduce the broader term of such a descriptor as a second indexing term thus creating a combination (see the 'Timbre' example above).

Coordination of terms and ambiguity

Apparently, single terms do not provide problems in indexing with descriptors derived from the UDC tables. However, the context can be very important in some

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

situations when disambiguation is necessary. Descriptors like Future - 81'366.584 or Voice - 81'366.57 can generate confusion in a universal context. That is why an extra descriptor is necessary to bring about the specification of the meaning or to add context to the mentioned descriptors. Even though the indexers would know what they deal with having the corresponding UDC notation next to the descriptor, the users will be lost in confusion about the specific meaning of these two descriptors. Consequently, some other descriptors will disambiguate the meaning of 'Future and Voice', namely 'Tense and aspect' in the former case and 'Grammatical categories' in the latter. It is the contextual disambiguation that will clarify the meanings of such descriptors.

On the other hand, ambiguity here affects only the terms of two of the languages involved. Romanian has a special word with only one meaning for this concept. Compare: Voice, Voix and Diateza.

Another related ambiguous situation arises for descriptors like: Revues, Revues, Reviste - 82-224, meaning a special kind of theatre performance. But, in French and Romanian the same term is used to denote a kind of periodical publication, an illustrated magazine. Therefore, a scope note is needed to specify or disambiguate the meaning of the descriptor.

Some precoordination was included in the thesaurus terms following the structure of the UDC notations. Consider the example:

SWISS LITERATURE - 821(494)
NT: French literature (Switzerland) - 821.133.1(494)
German literature (Switzerland) - 821.112.2(494)
Italian literature (Switzerland) - 821.131.1(494)

For the literature of multilingual countries like this, the solution adopted was to mention the name of the country in brackets for documents referring to literature in a particular language of that country. When the whole literary output of that country is considered then the appropriate descriptor will be Swiss literature. The same will be true for Canada and Belgium.

Literatures of ethnic minorities in Romania were treated the same way:

GERMAN LITERATURE (ROMANIA) - 821.112.2(498)
HUNGARIAN LITERATURE (ROMANIA) - 821.511.141(498)

In spite of its inherent controversy, this solution was adopted in order to confer greater specificity to the thesaurus, given the coordination of terms. The qualifier meant to modify the meaning of the descriptor allows the distinction of the same entity (an individual literature), but in this case, in a national context.

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

Brackets are used in another situation still. That is for the specification of the criteria by which some entities are grouped

e.g. DICTIONARIES (CONTENT)
DICTIONARIES (LANGUAGE)
DICTIONARIES (ORDERING)

An alternative solution here would have been a more complex descriptor formula such as:

DICTIONARIES ACCORDING TO CONTENT
DICTIONARIES ACCORDING TO LANGUAGE
DICTIONARIES ACCORDING TO ORDERING

but we found the solution of the qualifier as more convenient. The entire description of the notations for different categories of dictionaries was kept as scope notes.

Coordination can be substantival-adjectival (e.g. Children literature, Explanatory dictionaries), genitival (e.g. Word cohesion, Programming languages) and prepositional (Systems of orthography, Names of places).

Methodological principles

As mentioned above, the guiding rules for building the thesauri were provided by the international standards. We have applied them in keeping with the basic principles of thesauri construction and added some other specific rules to meet our goal.

(1) *The semantic factoring of the description of the UDC numbers* in the schedules or the combinations they may result in:
e.g. 811.111'36 is the result of a double synthesis, i.e. once the common auxiliary of language =111 for English was added to the main number for Linguistics 81, more precisely to 811, the "prefix" for an individual language, and a second time by adding the special auxiliary '36 meaning Grammar. This will be converted into descriptors as:

ENGLISH LANGUAGE + GRAMMAR
UF: English grammar

Another example: 81'276.3 Language or idiom of a particular age group or sex - was factored and rephrased in two descriptors:

LANGUAGE ACCORDING TO AGE
LANGUAGE ACCORDING TO SEX

(2) The selection of *the most commonly applied term from the UDC text as descriptor (preferred term),* leaving any synonyms as non-descriptors. There are different ways of representation of synonyms and near synonyms in UDC. They are separated by dot (.) as in: 81'23 Psycholinguistics. Psychology of language - or by comma (,) as in: 81-24 Dead, extinct languages. In both cases, a decision has to be made on the preferred form to become a descriptor, the other one being referred to as non-descriptor.

(3) *Keep the logical model for hierarchical arrangement* provided by the classification structure as far as possible. Sometimes the hierarchies in the UDC tables may be troublesome because of the decimal structure it is based on. An example of such a problem is found in the subdivision for 'Literary forms and genres' where the range of special auxiliaries exceed the number of ten and thus a formulation like 'Other literary genres' was imposed. This had consequences on the derived descriptors, therefore all the literary genres had to be mentioned in only one sequence.

(4) Preference for *terms proper to each of the three languages* to define certain concepts rather than the Latin forms (i.e. 'Nouns of place', 'Nouns of time' and 'Coordination' instead of 'Nomina loci', 'Nomina temporis', 'Parataxis'). This is a different situation from Biology where Latin is recommendable, as it has the position of 'lingua franca' unifying denominations of plants and animals.

(5) Preference for *plural forms* in concrete entities expressed by countable nouns (e.g. Poems, Novels, Adverbs, Literary genres). Abstract nouns are given in the singular (e.g. Morphology, Symbolism, Literary criticism).

(6) *Distinguish between singular and plural forms* in order to denote species and forms. Possible ambiguity of terms is eliminated by the scope notes and the relations each descriptor is provided with.

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

RONDEL
F: Rondel
R Rondel
UDC: 801.675.1
RT: French verse forms
Rondels

RONDELS
F: Rondeaux
R: Rondeluri
SN: Used to denote the
literary genre
UDC: 82-193.4
TG: Short poetic forms
RT: Rondel

Mention should be made here on another situation where distinction between singular and plural has another reason; disambiguation is necessary for only two of the three languages, English and French, where the terms are homonyms:

RECITATION
F: Rècitation
R: Recitare
UDC: 808.54
UF: Art of the disseur
BT: Rhetoric of speech

RECITATIONS
F: Rècitations
R: Recitaluri
UDC: 82-27
SN: Plays for a single performer
NT: Reading aloud
Art of story telling
RT: Monologues

(7) *Enrich the vocabulary of the UDC tables* with new terms used or to be used by the searchers. This is a case of very much used search terms like 'Poetic art' and 'Narrative art' for which we adopted a combination of descriptors corresponding to a combination of UDC numbers. The result was:

Poetic art
see: ART OF WRITING + POETRY
and
Narrative art
see: ART OF WRITING + PROSE

In these cases all the terms of the combinations have their corresponding UDC notation.

ISSN: 2324-9773

Francu, V. (1999). A universal classification system going through changes. Proceedings of the 10th ASIS SIG/CR Classification Research Workshop, 65-86. doi: 10.7152/acro.v10i1.12482

(8) *Use of the definition given by the UDC text as 'scope note'* where this is possible.
e.g.

81'282.4 National dialects. National variants or dialects of a language outside the country of origin

becomes:

NATIONAL DIALECTS
F: Dialects nationaux
R: Dialecte nationale
SN: Used for national variants or dialects of a language outside the country of origin
UDC: 81'282.4
BT: Dialects

(9) *Upward posting for narrower concepts given after* 'including' *in the UDC text.*
The UDC notation and its corresponding text:

81'367.52 Word arrangement. Positioning of parts of speech within clause.
Including: logical order, inverted order of words

This is expressed in the thesaurus in the following way:

Inverted word order
see: WORD ARRANGEMENT

Logical word order
see: WORD ARRANGEMENT

WORD ARRANGEMENT
F: Disposition des mots
R: Sintaxa propozitiei
SN: Used for the positioning of parts of speech within clauses
UDC: 81'367.52
BT: Syntax
UF: Inverted word order
Logical word order

The effect of this device is that specific terms are as good access points as the broader terms they are represented by.

Translatability

As mentioned earlier, the process of information transfer involves translation aspects. Translatability among terms belonging to different languages can be approached in

ISSN: 2324-9773

terms of semantics - i.e. equivalence or similarity regarding the meaning of terms, and in terms of morphology and syntax - i.e. equivalence or similarity regarding the structure of words and phrases.

Considering the semantic aspect of translation, there are various degrees of equivalence between the source language and the target language, representing the extent to which the meaning of a word in a language A is transferred to a word in another language B without loss of meaning. A task of primary importance in multilingual thesaurus construction is to always keep in mind the issue of language equality (Houdon, 1997).

Austin and Waters (1980) as much as Aitchison and Gilchrist (1987) identify several types of equivalence between terms of two or more languages:

1 exact equivalence, in the case of real cross-language or interlingual synonyms
e.g. ADVERBS / ADVERBES / ADVERBE
PROSODIC LICENCE / LICENCE PROSODIQUE / LICENTA PROZODICA

2 inexact equivalence, in the case of near synonyms, the solution suggested being a loan term from another language

e.g. WORD ARRANGEMENT / DISPOSITION DES MOTS / SINTAXA PROPOZITIEI
MACHINE TRANSLATION / TRADUCTION AUTOMATIQUE / TRADUCERE MECANICA

3 partial equivalence, also in the case of near synonyms, the solution being the use of broader or narrower terms in meaning

e.g. COMPUTATIONAL LINGUISTICS
LANGAGE DES ORDINATEURS
LINGVISTICA COMPUTATIONALA

4 single to multiple equivalence, the most complex situation requiring the most thoughtful analysis

e.g. CHILDREN LITERATURE / LITTÉRATURE POUR ENFANTS / LITERATURA PENTRU COPII
JUVENILE LITERATURE / ? / ?
Solution:
Juvenile literature
see: CHILDREN LITERATURE

ISSN: 2324-9773

5 non-equivalence in which case the use of loan terms is again recommended
e.g. AROMANIAN LITERATURE / LITTÉRATURE AROUMAIN / LITERATURA AROMÂNA
where the English and French terms are linguistic calques from Romanian.

As far as the morphological and syntactic forms of the descriptors are concerned, we can say that language names make a special category of terms in the thesaurus to illustrate this. Most of them are *cognates* (pairs of words having semantic and formal identity) in all three languages (e.g. Aymara, Dakota, Bengali). Not only in language names can we find cognates but also in other terms (Adverb, Apocope, Dialect). *Phonological identity* (Aztec / Aztéque / Azteca or Yiddish / Yiddish / Idis) and *phonetic alternances* mainly in word endings are frequently met:

e.g.
ESTONIAN / ESTONIEN / ESTONIANA
ILLYRIAN / ILLYRIEN / ILIRA
LIGURIAN / LIGURIEN / LIGURICA

We may just as well have semantic identity with *no formal identity* at all as in:

DUTCH / NÉERLANDAIS / OLANDEZA
SPEECH ANALYSIS / ANALYSE DU DISCOURS / ANALIZA VORBIRII

Conclusion

Gilchrist (1992) makes a statement by which he pronounces UDC as an improvement on DDC and BC2 as an improvement on both. For all that, DDC is more widely used than UDC, whereas UDC is more widely used than BC2. His conclusion is given as the following question: "is it unconceivable to think of some fruitful collaboration between these three classification schemes?"

Along similar lines, Scibor (1997) predicts that a UDC-based thesaurus will only be feasible if the structure of the entire classification system is changed into a fully faceted scheme based on postcoordination. Remarkable progress in restructuring the UDC has been achieved by McIlwaine and Williamson (1997) in Class 61 - Medical Sciences, since the beginning of their pilot project in 1993. The restructuring of Class 61 has been made in keeping with the faceted structure of Class H for Health Sciences, Biomedical Sciences in Bliss Bibliographic Classification, second edition (BC2).

The present project is another proof that the UDC schedules are worth more attention and re-evaluation of such an important resource as a universal classification system

ISSN: 2324-9773