

# Introduction

1. Purpose *ix*
2. History *x*
3. Principles of the new edition *x*
4. Subject coverage *x*
5. Source of terms *xi*
6. Structure of the Thesaurus *xi*  
*Terms*  
*Relationships between terms*
7. Presentation *xiii*  
*Part 1: Alphabetical structured and permuted list*  
*Part 2: Hierarchical list by microthesaurus*  
*Part 3: French/English/Spanish index of descriptors*  
*Part 4: Spanish/English/French index of descriptors*
8. Characteristics *xv*
9. Conventions *xv*
10. General guidelines for the use of the Thesaurus *xvi*  
*Guidelines for indexing documents*  
*Guidelines for formulating queries*

## 1. PURPOSE

The UNESCO Thesaurus is designed to be the main working tool of the UNESCO Computerized Documentation System (CDS). It is a trilingual controlled vocabulary principally used in indexing and retrieval of information stored in the UNESCO Bibliographic Database (UNESBIB), and various other sub-databases which constitute the UNESCO Integrated Documentation Network. The material analyzed includes the following:

- UNESCO documents:
  - Field mission reports,
  - Working series and conference documents,
  - Main series documents,
  - General Conference documents,
  - Executive Board documents,
- UNESCO publications,
- UNESCO periodicals,
- Regional Office and UNESCO Institute documents and publications,
- UNESCO-sponsored publications,
- UNESCO Library acquisitions,
- Acquisitions of the various Divisions,
- Collections of the various specialized UNESCO documentation centres.

Thesaurus terms are used in the bibliographic publications produced by the UNESCO Library and Documentation Service, namely:

- UNESCO List of Documents and Publications (ULDP), issued quarterly, with annual cumulations (available in print, on CD-ROM and, online, on the ECHO and IDRC hosts),
- UNESCO Library Acquisitions (ULA), issued quarterly,
- Annual and multi-annual indexes to leading UNESCO periodicals, e.g. *Impact of Science on Society*; *International Social Science Journal*; *Museum International*; *Nature and Resources*; *Prospects: a quarterly review of education and the UNESCO Courier*,
- Indexes to field mission reports, Director-General's speeches, etc.

In addition to these publications, terms from the Thesaurus are also used regularly to retrieve information on user request from UNESBIB, either in retrospective bibliographic searches (batch or on-line) or in SDI (Selective Dissemination of Information) searches.

Users of the UNESCO Thesaurus are all institutions and individuals interested in accessing UNESCO-produced documentary information:

- Institutions in Member States, including National Commissions,
- United Nations System and other intergovernmental organizations,
- International non-governmental organizations,
- Experts and consultants,
- UNESCO staff and visitors to the Organization.

The UNESCO Thesaurus may also serve as a reference tool for other UNESCO information services, for documentation centres in Regional Offices, affiliated non-governmental organizations and information/library services in Member States, as they develop their own specialized information processing vocabularies and systems.

## **2. HISTORY**

The UNESCO Computerized Documentation System (CDS) was developed during 1969-72 and became operational in 1973. Before indexing started, an unstructured list of descriptors was compiled of terms selected from the OECD *Macrothesaurus* and from basic UNESCO documents. Over the years, terms were added and others deleted, thus reflecting the experience gained through regular indexing of the literature.

This list of terms formed the core of the draft edition of the UNESCO Thesaurus, which was compiled between April 1973 and April 1974. The draft thesaurus of 5,000 terms was then distributed to professional staff and specialists within UNESCO for comment, while the Indexing and Retrieval Section of the Documentation Systems Division started to use it as a source of new index terms, on an experimental basis. It was found necessary to make considerable extensions to certain subject areas, as the result of comments and recommendations received from UNESCO programme divisions and experts. Many new terms were added bringing the total to approximately 11,500 entries.

The first edition of the UNESCO Thesaurus was released in English in 1977, with French and Spanish translations in 1983 and 1984.

But to keep abreast of developments in the documents analysed, in the light of the evolving priorities of UNESCO programmes, and of the needs of a growing number and variety of users, it became necessary to prepare a revised, updated and restructured edition of the Thesaurus.

## **3. PRINCIPLES OF THE NEW EDITION**

The first stage of this revision, initiated in-house by systematic and continuous updating and maintenance of both the Thesaurus and the bibliographic databases, involved the addition of new terms (descriptors and non-descrip-

tors), deletion of those never used, retention of those lesser used by leading them into preferred terms; increase of the number of explanatory scope notes, revision of associative and hierarchical relationships, etc. Certain subject areas were expanded (development, environment, information); others, especially science and technology, had to be compressed so as to reflect policy, rather than pure research orientations of the documentary material analyzed. Priority was given to the Education field, in the light of its terminological harmonization with the UNESCO:IBE Thesaurus. As a result, the number of terms (descriptors and non-descriptors) decreased to some 6,600.

In the second stage, a prototype thesaurus database was set up. The systematic display of terms was restructured according to the Microthesaurus approach, in which subject areas are displayed in such a way so as to allow the user an easy overview of a particular area at a glance. This should significantly enhance the userfriendliness of the Thesaurus.

With the same objective in mind, a permuted list of descriptors and non-descriptors was created, which allows to access the compound terms by each of their significant words, alphabetically sorted. It was also decided to merge this permuted list with the alphabetical structured list, which is usually sorted only by the first word of the descriptors and non-descriptors: the display thus obtained permits the user to find all alphabetical entries, whether on the first words or on any other significant words, in a single display.

Finally, this revised version includes the translation of all descriptors into French and Spanish, and two trilingual indexes, one sorted by French descriptors, and the other, by Spanish descriptors.

These new presentations (the alphabetical structured and permuted display, the subject-oriented hierarchical display by microthesaurus and the trilingual indexes) required a new design of the Mainframe CDS/ISIS software for Thesaurus management.

## **4. SUBJECT COVERAGE**

The Thesaurus covers the major fields of knowledge that constitute the scope of UNESCO; within each field, a series of microthesauri regroup the terms according to their meaning. Here is the list of fields and microthesauri.

### **1. Education**

- 1.05 Educational sciences and environment**
- 1.10 Educational policy**
- 1.15 Educational planning**
- 1.20 Educational administration**
- 1.25 Educational management**
- 1.30 Educational systems and levels**
- 1.35 Educational institutions**
- 1.40 Curriculum**
- 1.45 Basic and general study subjects**
- 1.50 Technical and vocational study subjects**
- 1.55 Educational population**

- 1.60 Teaching and training
  - 1.65 Educational evaluation
  - 1.70 Educational facilities
- 2. Science**
- 2.05 Scientific approach
  - 2.10 Science and research management
  - 2.15 Mathematics and statistics
  - 2.20 Physical sciences
  - 2.25 Chemical sciences
  - 2.30 Space sciences
  - 2.35 Earth sciences
  - 2.40 Geography and oceanography
  - 2.45 Meteorology
  - 2.50 Hydrology
  - 2.55 Environmental sciences and engineering
  - 2.60 Pollution, disasters and safety
  - 2.65 Natural resources
  - 2.70 Natural sciences
  - 2.75 Medical sciences
  - 2.80 Pathology
- 3. Culture**
- 3.05 Culture
  - 3.10 Cultural policy and planning
  - 3.15 Philosophy and ethics
  - 3.20 Religion
  - 3.25 History
  - 3.30 Linguistics
  - 3.35 Languages
  - 3.40 Literature
  - 3.45 Arts
  - 3.50 Visual arts
  - 3.55 Performing arts
  - 3.60 Museums
  - 3.65 Leisure
- 4. Social and human sciences**
- 4.05 Social sciences
  - 4.10 Psychology
  - 4.15 Social systems
  - 4.20 Social problems
  - 4.25 Social welfare and policy
  - 4.30 Ethnic questions
  - 4.35 Population
  - 4.40 Family
  - 4.45 Human settlements and land use
- 5. Information and communication**
- 5.05 Information sciences
  - 5.10 Communication research and policy
  - 5.15 Information management
  - 5.20 Information industry
  - 5.25 Documentary information systems
  - 5.30 Information sources
  - 5.35 Documentary information processing
  - 5.40 Information technology (software)
  - 5.45 Information technology (hardware)

- 6. Politics, law and economics**
- 6.05 Legal systems
  - 6.10 Human rights
  - 6.15 Politics and government
  - 6.20 International relations
  - 6.25 Economics
  - 6.30 Economic and social development
  - 6.35 Agriculture
  - 6.40 Industry
  - 6.45 Civil, military and extraction engineering
  - 6.50 Manufacturing and transport engineering
  - 6.55 Materials and products
  - 6.60 Equipment and facilities
  - 6.65 Services
  - 6.70 Finance and trade
  - 6.75 Organization and management
  - 6.80 Personnel management
  - 6.85 Labour
- 7. Countries and country groupings**
- 7.05 Africa
  - 7.10 America and the Caribbean
  - 7.15 Asia and the Pacific
  - 7.20 Europe
  - 7.25 Political groupings
  - 7.30 Economic groupings
  - 7.35 Geographic groupings
  - 7.40 Ethnic and religious groupings
  - 7.45 Linguistic groupings

## 5. SOURCE OF TERMS

The main source for the second edition was the UNESBIB database, which permitted to display for each descriptor its frequency of occurrence in document indexing and, when necessary, its co-occurrence with other descriptors and a sample of the documents indexed.

In case of doubt, the last version of the OECD *Macrothesaurus* was systematically consulted. More specialized thesauri were also examined, such as the EUDISED Thesaurus in the field of education, the SPINES Thesaurus in the field of science, the ILO Thesaurus for labour and occupational life, etc., with a view to enhancing terminological compatibility with other international controlled vocabularies.

## 6. STRUCTURE OF THE THESAURUS

The UNESCO thesaurus contains:

- terms,
- relationships between terms.

### 6.1 Terms

#### *Field names*

A field is a grouping of microthesauri; a field name is preceded by a one-digit serial number, used to obtain the same sorting of the fields in each language of the Thesaurus. A field name is not necessarily a descriptor;

Example:

## 5. Information and communication

### *Microthesaurus names*

A microthesaurus is a grouping of descriptors and non-descriptors, designed in such a way that its content holds on no more than one page, so that the user can browse it at a glance; a microthesaurus name is preceded by a three-digit serial number, used to obtain the same sorting of the microthesauri in each language of the Thesaurus, and to refer each descriptor in the alphabetical list of the Thesaurus to the microthesaurus to which it belongs; the first of the three digits is the number of the subject field to which the microthesaurus belongs. A microthesaurus name is not necessarily a descriptor;

Example:

### 1.45 Basic and general study subjects

### *Descriptors*

A descriptor, also called preferred term, is the term to be used to represent a concept when indexing documents and formulating queries; the number of descriptors is the same in the three languages of the Thesaurus;

Example:

### Sustainable development

### *Non-descriptors*

A non-descriptor, also called forbidden term, is a form designating a concept very close to that represented by a descriptor; it is, as the corresponding descriptor, an access point in the Thesaurus, whereby the free language used by the authors of documents and by the user making a search is directly towards the standardized terminology of the descriptors. Since the number of non-descriptors may vary from one language to another, it was decided to assign non-descriptors only to English descriptors;

Example:

### *Ecodevelopment*

### *Scope notes*

A scope note, shown by the symbol **SN**, briefly explains the use for which a descriptor is intended when its meaning is ambiguous;

Example:

### Education and culture

**SN** Taking into consideration of cultural factors in the educational process.

## 6.2 Relationships between terms

### *Inclusion*

Relation between a descriptor and the microthesaurus to which it belongs, shown by the symbol **MT** (= Microthesaurus), placed between a descriptor and a microthesaurus number and name;

Example:

### Sustainable development

**MT** 6.35 Economic and social development

### *Hierarchy*

Relation between descriptors belonging to the same semantic arborescence, shown by the symbols:

**BT** (= Broader Term, i.e. generic term), placed between a specific descriptor and a generic descriptor;

Example:

### Sustainable development

**BT1** Development strategies

**NT** (= Narrower Term, i.e. specific term), placed between a generic descriptor and a specific descriptor;

Example:

### Development strategies

**NT1** Sustainable development

Remarks:

1) The digit to the right of the symbol **BT** or **NT** indicates the number of hierarchical levels separating the descriptors of an arborescence;

Example:

### Chronology

**BT1** Historical method

**BT2** History

**NT1** Dating

**NT2** Archaeological dating

**NT2** Radiocarbon dating

2) A descriptor that has no broader term is called a top term.

3) The hierarchical relationships are identical in all the language versions of the Thesaurus.

4) The multihierarchy, i.e. the fact that a descriptor can belong to more than one microthesaurus, was adopted only for country names (in field n° 7): a country is indeed always included in a continent, and can also belong to one or several political, economic, geographic, ethnic, religious and linguistic groupings.

### *Association*

Relation between two descriptors linked by an association of ideas is shown by the symbol **RT** (= Related Term, i.e. associated term), placed between two associated descriptors;

Examples:

### Environmental management

**RT** Sustainable development

...

### Sustainable development

**RT** Environmental management

Remarks:

1) The associative relationships are identical in all the language versions of the Thesaurus.

2) Two descriptors cannot be linked by both a hierarchic and an associative relationship.

3) The associative relationship between descriptors sharing at least one significant word has usually not been created, in order not to expand uselessly the volume of the Thesaurus: indeed, these descriptors appear in sequence in

the alphabetical list, either as normal entries, or as permuted entries, thus permitting the user to establish himself the relationship by glancing through neighbouring entries;

Example 1:

**Disarmament** (as a descriptor entry)

**Disarmament education** (as the next descriptor entry)

Example 2:

**Curriculum** (as a descriptor entry)

*Curriculum:* (as a significant word entry)

**Integrated curriculum** (as a descriptor including the word: curriculum)

For a more detailed presentation of permuted entries, see § 7.1., *significant word entry*, hereafter.

### *Semantic equivalence*

Relation between a descriptor and a non-descriptor, shown by the symbols:

**USE**, placed between a non-descriptor and the corresponding descriptor;

Example:

*Ecodevelopment*

**USE Sustainable development**

**UF** (= Used For), placed between a descriptor and a non-descriptor which it represents;

Example:

**Sustainable development**

**UF** *Ecodevelopment*

Remarks:

- 1) A non-descriptor must refer to one descriptor only.
- 2) A descriptor may have zero, one, two or more non-descriptors corresponding to it.

### *Linguistic equivalence*

Relation between descriptors designating the same concept in different languages, shown by the symbol of the language indicators:

**FR** (= French), placed between an English descriptor and the corresponding French descriptor;

**SP** (= Spanish), placed between an English descriptor and the corresponding Spanish descriptor;

Example:

**Sustainable development**

**FR** *Développement durable*

**SP** *Desarrollo durable*

## 7. PRESENTATION

The UNESCO Thesaurus is published in one volume consisting of four parts, in addition to the introduction:

Part 1: Alphabetical structured and permuted list sorted by English terms, with their French and Spanish equivalents.

Part 2: Hierarchical list by microthesaurus.

Part 3: French/English/Spanish index of descriptors.

Part 4: Spanish/English/French index of descriptors.

## 7.1 Part 1: Alphabetical structured and permuted list

The alphabetical structured and permuted list includes three types of alphabetically sorted entries: descriptors, non-descriptors, significant words other than the first word of a descriptor or of a non-descriptor; the content of each type of entry is described hereafter.

### *Descriptor entry*

- Wording of the descriptor, in bold,
- Number and name of the microthesaurus to which the entry descriptor belongs, preceded by the symbol **MT**,
- Language equivalent in French, in italics, preceded by the language indicator **FR**,
- Language equivalent in Spanish, in italics, preceded by the language indicator **SP**,
- Scope note, if present, preceded by the symbol **SN**,
- Non-descriptor(s) corresponding to the entry descriptor, when it has at least one, in italics, preceded by the symbol **UF** (= Used For), and sorted alphabetically,
- Generic descriptors of the entry descriptor, if it has at least one, in bold, preceded by the symbol **BT** (= Broader Term) and a digit indicating the number of hierarchical levels between the entry descriptor and each generic descriptor; the generic descriptors are sorted:
  - in order of ascending hierarchical levels,
  - within each hierarchical level, in alphabetical order,
- Specific descriptor(s) of the entry descriptor, if it has at least one, in bold, preceded by the symbol **NT** (= Narrower Term) and a digit indicating the number of hierarchical levels between the entry descriptor and each specific descriptor; the specific descriptors are sorted:
  - in order of descending hierarchical levels,
  - within each hierarchical level, in alphabetical order,
- Descriptor(s) associated with the entry descriptor, if it has at least one, in bold, preceded by the symbol **RT** (= Related Term), and sorted alphabetically;

Example:

**Universal education**

**MT** 1.10 Educational policy

**FR** *Éducation universelle*

**SP** *Educación universal*

**SN** System of education extending opportunities to all.

**UF** *Equal education*

**UF** *Equal opportunity (education)*

**BT1** **Educational opportunities**

**BT2** **Right to education**

**NT1** **Compulsory education**

**NT1** **Free education**

**RT** **Access to education**

**RT** **Democratization of education**

**RT** **Educational discrimination**

**RT** **Educationally disadvantaged**

### *Non-descriptor entry*

- Wording of the non-descriptor, in italics,
- Wording of the corresponding descriptor, in bold, preceded by the symbol **USE**, and followed by the num-

ber, in brackets, of the microthesaurus of that descriptor;

Example:

*Bird migration*

**USE Animal migration (2.70)**

### *Significant word entry*

- Wording of a significant word (uniterm) of a descriptor or of a non-descriptor (other than the first word which already appears in the alphabetical sorting of the descriptor and non-descriptor entries), in italics, followed by a colon;
- Sub-entry that includes a second, third, etc., significant word of a descriptor:
  - complete wording of the descriptor, in bold, followed by the number, in brackets, of the microthesaurus of that descriptor,
- Sub-entry that includes the second, third, etc., significant word of a non-descriptor:
  - complete wording of the non-descriptor, in italics, wording of the corresponding descriptor, in bold, preceded by the symbol **USE**, and followed, by the number, in brackets, of the microthesaurus of that descriptor,
- All sub-entries are alphabetically sorted;

Example:

*Change:*

**Attitude change (4.10)**

*Biological change*

**USE Mutation (2.75)**

*Career change*

**USE Occupational mobility (6.75)**

**Climatic change (2.65)**

**Cultural change (3.05)**

*Educational change*

**USE Educational reform (1.10)**

**Language change (3.30)**

*Occupational change*

**USE Occupational mobility (6.75)**

**Organizational change (6.70)**

*Population change*

**USE Population dynamics (4.35)**

**Resistance to change (4.15)**

**Social change (4.15)**

**Sociology of change (4.05)**

**Technological change (2.10)**

## **7.2 Part 2: Hierarchical list by microthesaurus**

The hierarchical list by microthesaurus includes one page by microthesaurus, which is identified by the number and name of the microthesaurus (centered at the upper part of the page), and sorted by microthesaurus number.

For each microthesaurus, the entry terms are the top term descriptors (i.e. the descriptors without any broader term), in bold, sorted alphabetically.

Beneath each entry descriptor, one finds:

- the non-descriptor(s), if any, corresponding to the top term descriptor, in italics, preceded by the symbol **UF** (= Used For), and sorted alphabetically,
- the descending hierarchy of descriptors constituting

the arborescence of the top term descriptor; each descriptor, in bold, is preceded by the symbol **NT** (= Narrower Term) and a digit indicating the number of hierarchical levels between the top term descriptor and each specific descriptor; this number of hierarchical levels is also marked by an indentation of five characters by hierarchical level after the first one; the specific descriptors are sorted:

in order of descending hierarchical levels,

within each hierarchical level, in alphabetical order,

- if a descriptor has one or several non-descriptor(s), these non-descriptors are displayed, in italics, and in alphabetical order, beneath the descriptor;

Example:

### 3.05 CULTURE

<b>Cultural anthropology</b>	<b>NT1</b> Countercultures
<i>UF Cultural and social anthropology</i>	<b>NT1</b> Cultural pluralism
	<i>UF Cultural diversity</i>
	<b>NT2</b> Biculturalism
	<b>NT2</b> Multiculturalism
<b>Cultural change</b>	<b>NT1</b> Culture of poverty
<i>UF Cultural mutation</i>	<b>NT1</b> Culture of work
<b>NT1</b> Civilization crises	<i>UF Culture of enterprise</i>
<b>NT2</b> Cultural crises	<b>NT1</b> Disappearing cultures
<b>NT1</b> Cultural dynamics	<b>NT1</b> Dominant cultures
<b>NT1</b> Cultural innovations	<i>UF Cultural dominance</i>
<b>NT1</b> Cultural revolution	<i>UF Cultural imperialism</i>
	<b>NT1</b> Mass culture
<b>Cultural conditions</b>	<i>UF Popular culture</i>
<i>UF Cultural situation</i>	<b>NT1</b> Scientific culture
<b>NT1</b> Cultural environment	<i>UF Technical culture</i>
<b>NT1</b> Cultural factors	<b>NT1</b> Subcultures
<b>NT1</b> Cultural life	<b>NT1</b> Traditional cultures
<b>NT2</b> Everyday life	<b>NT1</b> Universal culture
<b>NT2</b> Life styles	<b>NT1</b> Working class cultures
<b>Cultural creation</b>	
	<b>Culture</b>
<b>Cultural differentiation</b>	<b>NT1</b> Civilization
<b>NT1</b> Cultural discrimination	<b>NT1</b> Cultural history
<b>NT1</b> Cultural elite	<b>NT1</b> Cultural philosophy
<b>NT1</b> Cultural inequality	<i>UF Concept of culture</i>
<b>NT1</b> Cultural minorities	<b>NT1</b> Cultural sociology
<i>UF Minority cultures</i>	
<b>Cultural identity</b>	<b>National cultures</b>
<i>UF Cultural alienation</i>	<i>UF Cultures</i>
<i>UF Cultural identification</i>	<b>NT1</b> African cultures
<i>UF Ethnic identity</i>	<b>NT1</b> Amerindian cultures
<b>NT1</b> Acculturation	<i>UF North American Indian cultures</i>
<i>UF Cultural assimilation</i>	<i>UF South American Indian cultures</i>
<i>UF Cultural integration</i>	<b>NT2</b> Aztecs
<i>UF Enculturation</i>	<b>NT2</b> Incas
<b>NT1</b> Cultural nationalism	<b>NT2</b> Mayas
<b>NT2</b> National identity	<b>NT1</b> Arab culture
<i>UF National character</i>	<b>NT1</b> Arctic cultures
<i>UF National consciousness</i>	<b>NT1</b> Asian cultures
	<i>UF Oriental cultures</i>
<b>Cultural interaction</b>	<b>NT2</b> Central Asian cultures
<i>UF Cultural contact</i>	<i>UF Malay culture</i>
<i>UF Cultural influence</i>	<i>UF Mongolian culture</i>
<b>NT1</b> Cultural conflicts	<i>UF Tamil cultures</i>
<b>NT1</b> Cultural isolation	<b>NT1</b> Caribbean cultures
	<b>NT1</b> European cultures
<b>Cultural research</b>	<i>UF Western cultures</i>
<b>NT1</b> Cross cultural analysis	<b>NT2</b> Baltic cultures
<b>NT1</b> Cultural models	<b>NT2</b> Central European cultures
<b>NT1</b> Cultural studies	<b>NT2</b> Iberian cultures
	<b>NT2</b> Slav cultures
<b>Cultural systems</b>	<b>NT1</b> Latin American cultures
<i>UF Cultural types</i>	<b>NT1</b> North American cultures
<b>NT1</b> Contemporary culture	<b>NT1</b> Oceanic cultures

## **7.3 Part 3: French/English/Spanish index of descriptors**

The multilingual index, sorted by French descriptor, is presented in three columns:

- in the first column, all French descriptors of the Thesaurus are listed, in alphabetical order, and in bold,
- in the second column, the corresponding English descriptors are displayed, in normal character types.

- in the third column, the corresponding Spanish descriptors are displayed, in normal character types.

**Remark:**

The non-descriptors do not appear in the multilingual index, as there is no systematic correspondance between the non-descriptors of the three language versions of the Thesaurus;

**Example:**

FRANÇAIS	ANGLAIS	ESPAGNOL
<b>Auberge de jeunesse</b>	Youth hostels	Albergue juvenil
<b>Autoassistance</b>	Self help	Esfuerzo propio
<b>Autoenseignement</b>	Self instruction	Autoaprendizaje
<b>Autogestion</b>	Self government	Autogestion
<b>Automatation</b>	Automation	Automatizacion
<b>Autonomie éducative</b>	Educational autonomy	Autonomia educativa

**7.4 Part 4:  
Spanish/English/French index  
of descriptors**

The multilingual index, sorted by Spanish descriptor, is presented in three columns:

- in the first column, all the Spanish descriptors of the Thesaurus are listed in alphabetical order, and in bold,
- in the second column, the corresponding English descriptors are displayed, in normal character types,
- in the third column, the corresponding French descriptors are displayed, in normal character types.

**Remark:**

There is no need for an index sorted by English descriptors, as these descriptors are sorted alphabetically in Part I of the Thesaurus;

**Example:**

ESPAÑOL	INGLÉS	FRANÇÉS
<b>Animal domestico</b>	Domestic animals	Animal domestique
<b>Animal marino</b>	Marineanimals	Animal marin
<b>Animal salvaje</b>	Wild animals	Animal sauvage
<b>Animismo</b>	Animism	Animisme
<b>Año académico</b>	Academic year	Année académique
<b>Antecedentes académicos</b>	Educational background	Acquis éducatif

**8. CHARACTERISTICS**

All the language versions of the UNESCO Thesaurus comprise:

- 87 microthesauri,
- 4261 descriptors,
- 4049 reciprocal hierarchical relations (**BT/NT**),
- 5771 associative relations (**RT**).

The English version includes 2341 non-descriptors and 601 scope notes.

The Thesaurus is structured according to the ISO standards 2788 and 5964.

**9. CONVENTIONS**

*Number*

The EJC (Engineers Joint Council) rules are followed on singulars and plurals for the English terms. These rules specify that «in choosing between singular and plural noun forms, a useful rule of thumb may be applied as follows: use the plural form when the proposed term is a count noun, that is, a noun which may be used to answer the question «how many?» ...; use the singular form for mass nouns, those that express «How much?» ...; use the singular for specific processes, properties, or conditions. The French and Spanish terms are in the singular form.

*Oblique*

Occasionally, it has been necessary to include compound terms containing an oblique. This arose where there was no single term available to express the concept. For example, for library science and information science there is no general descriptor covering both disciplines; consequently it has been necessary to coin the term Information/library to precede administration, personnel and other aspects common to both sciences.

*Pre-coordinated terms*

The Thesaurus contains many pre-coordinated concepts, that is, compound terms made up of two or more words. The majority of these terms are combinations of two words, e.g. «Scientific institutions», «Sustainable development», etc . There are a few longer terms such as «Cultural property preservation», «Right to natural resources control», but these are not frequent.

*Alphabetization*

The word-by-word arrangement is used in alphabetization, for example:

- Social values**
  - Social welfare**
  - Social work**
  - Socialism**
  - Socialization**
- and oblique terms follow the term alone:
- Information users**
  - Information work**
  - Information/library administration**

**10. GENERAL GUIDELINES FOR THE USE OF THE THESAURUS**

The UNESCO Thesaurus may, as any other thesaurus, be used:

- by professional indexers, to index documents for feeding a bibliographic database,
- by professional librarians, documentalists and information officers, as well as by end-users, to formulate queries in order to search the bibliographic database.

## 10.1 Guidelines for indexing documents

Indexing consists in listing the concepts contained in a document and in representing those concepts by thesaurus descriptors.

Indexing is designed to make it possible to select from the set of bibliographical references held in a database the references matching the queries formulated. Selection is possible only if the concepts to which a query relates are expressed with the same thesaurus descriptors in both the query and in the relevant documents.

In practice, indexing proceeds as follows:

- a) The indexer briefly reads the content of the document to be indexed,
- b) The indexer then selects the main concepts of the document with due regard to the following two rules:
  - *selectivity*: only those concepts must be selected on which the document provides information which is likely to be of interest to users. If an indexer selects a concept just cited in a document, but without any useful information about it, the users of the database will find that document, although not relevant, while searching on that particular concept,
  - *exhaustiveness*: all useful concepts appearing in the text must be selected, whether they are referred to explicitly or implicitly; if an indexer disregards a concept about which a document brings useful information, the users will not retrieve that document, although it is relevant, while searching on that particular concept,
- c) The indexer then translates the selected concepts into the corresponding thesaurus descriptors starting with a search, in Part 1 (Alphabetical structured and permuted list) of the Thesaurus, for alphabetical entries corresponding to the wording of the concepts (if necessary, preliminarily translated into the language of the used language version of the Thesaurus ). Three cases may occur:
  - a descriptor exists which matches the concept in meaning; this is then entered on the indexing form or on-screen entry form (e.g. «Comprehension»),
  - a non-descriptor exists which matches the concept in meaning; the non-descriptor refers back to the descriptor to be used, and the procedure is the same as above (e.g. «Ecodevelopment», which refers back to «Sustainable development»),
  - no descriptor nor non-descriptor may be found in the Thesaurus, whose wording matches that of the concept in the document; in that case, the indexer should go forward according to one or several of the successive following options:
    - if the concept is expressed in a multi-term form, the indexer should explore the alphabetical list for a significant word entry that would match the second, third, etc. word of the concept and that would refer to a descriptor or non-descriptor with the same meaning as the concept (e.g. «demographic increase» is neither a descriptor nor a non-descriptor; but if one searches

«increase», one will find «Population increase» as a descriptor),

the indexer may use imagination, by using other words to express the concept and see whether there is a matching entry in the alphabetical list (e.g. «mainframe» is neither a descriptor, nor a non-descriptor, nor a significant word of a multi-term descriptor or non-descriptor; the indexer should think that a mainframe is a kind of computer; and then find that «Computers» is indeed a descriptor),

the indexer may select two or more descriptors in the alphabetical thesaurus to express the concept to be represented (e.g. «household budget» is not a descriptor; that concept may however be represented by the combination of the two conceptual components «budget» and «household», which may be expressed by the descriptors «Budgets» and «Households», as the same process will be performed when representing that concept during a search, the statement will be retrieved),

finally, the indexer may define the general category or categories (microthesauri) which encompasses or encompass the refractory concept in question, consult Part 2 (Hierarchical list by microthesaurus) of the Thesaurus and search among the available descriptors for those which most closely approximate that concept (e.g. the concept of «loss of the quality of earth» is not to be found in the Thesaurus; it however clearly comes under the «Pollution, disasters and safety» microthesaurus; the indexer will screen that microthesaurus, and will rapidly find the descriptor «Soil degradation», which is appropriate to represent the selected concept).

Two other rules are used to represent concepts by means of thesaurus descriptors:

- *vertical specificity*: the descriptor must be located at the same level of specificity as the concept itself or, failing this, at the next level up within the Thesaurus;  
Examples:  
A document concerning the concept of foreign language must be indexed by the descriptor «Foreign languages» and not by the generic descriptor «Languages».  
However, a document on influenza - a concept which is not included in the thesaurus - must be indexed by the broader descriptor «Infectious diseases».
- *horizontal specificity*: a concept made up of two or more words must be expressed by a compound descriptor if one exists, rather than by combining individual descriptors;  
Example:  
A document concerning computer science should be indexed by the descriptor «Computer science» and not by the descriptors «Computers» and «Science».

The interest of the specificity rules is that if documents



are indexed by descriptors with the same degree of specificity as the concepts which they express (e.g. «Foreign languages»), queries relating to those descriptors will generate a bibliographical reference list that contains no document indexed at a more generic level (e.g. «Languages») and will thus be very accurate.

On the contrary, if the documents are indexed at the level which is more generic than the concepts they contain, a search on the corresponding descriptors will bring a lot of non-relevant references.

In summary, proper application of the rules on selectivity, exhaustiveness and specificity is an essential factor determining the quality of a documentary database.

d) Finally, the few remaining concepts for which the indexer will not have found any appropriate descriptors in the Thesaurus, if really worthwhile, will be kept as such, as 'free' or 'candidate' descriptors. The UNESCO Thesaurus manager will screen these terms from time to time, and possibly include them in the Thesaurus during a thesaurus maintenance operation.

## **10.2 Guidelines for formulating queries**

Formulating a conceptual query is analogous to the indexing of a document. Indexing consists in listing the concepts to which a given document relates and on which information is provided that is likely to be of interest to users, the concepts being expressed by descriptors. The process of formulating a query consists in listing the concepts that are likely to be included in the documents and that are the subject of the question, in converting those concepts into descriptors and subsequently in conducting a search within the documentary database for the documents which have been indexed by those descriptors.

The listing of concepts for a query however calls for a very different approach from that taken when listing concepts in connection with a document; document indexing is a process which results in a very considerable degree of compression: a five to ten page document is usually represented by between five and fifteen descriptors (or a compression rate of between 100:1 and 1000:1) while the process of formulating a query usually increases the volume of the original query statement.

When formulating a query, users begin by setting out

in natural language, in as much detail as possible, the requested information; users list the concepts, which, when combined, constitute the substance of the question.

Users must then choose Thesaurus descriptors to represent those concepts; three cases may occur:

1) The query is precisely formulated and the concepts which it encompasses are included in the Thesaurus: users should preferably consult Part 1 (Alphabetical structured and permuted list) of the Thesaurus, where they will find:

- the descriptors or non-descriptors (and from them the descriptors) corresponding to the concepts in the query,
- the hierarchically and associatively related descriptors which can help them to broaden the scope of the query,

2) The question is precisely formulated, but the Thesaurus contains none or only some of the concepts which the query encompasses: after unsuccessfully searching in the alphabetical structured and permuted list, users should consult Part 2 (Hierarchical list by microthesaurus) of the Thesaurus, and especially the microthesaurus or microthesauri corresponding to the fields or field covered by the query concerned, and conduct a search for existing descriptors covering those concepts which are most closely approximate to those indicated in the query,

3) The question is not precisely formulated, which is a common occurrence: users should preferably consult the microthesaurus or microthesauri corresponding to the field or fields covered by the query; they will find there the hierarchical list of descriptors covering the field or fields concerned and can select the most appropriate descriptors.