PREFACE

As long ago as 1981, the Commission of the European Communities and the Council of Europe decided to develop a common documentary terminology in nine languages.

The *European Education Thesaurus*, published in 1991, represented the result of the efforts made by the two organisations over some ten years to complete and update the previous EUDISED Thesaurus by bringing its vocabulary into line with new developments in education.

This new edition, issued seven years later in 1998, illustrates the continuity of this cooperation. The common objective is to achieve the widest possible distribution of a regularly updated tool adapted to the needs of all those in Europe who are involved in the exchange of information on education.

The main features of this third edition are not only the inclusion of all the amendments accepted since 1991 by the group responsible for the management of the content of the *European Education Thesaurus*, working closely with the various language areas concerned, but also its extension to embrace Swedish and Finnish, the two new official languages of the European Union.

Various *European Education Thesaurus* versions in other languages have been developed in parallel – Albanian, Croatian, Czech, Hungarian, Polish, Slovenian and Turkish. These are proof of the interest that an increasing number of countries are taking in this tool and its use and adaptation at national level.

So as to make the *European Education Thesaurus* accessible as widely as possible to those interested in it, it has been decided that this new edition will be made available primarily on the Internet. The next stage of cooperation between the Commission and the Council of Europe will be to develop a permanent management and updating tool for it, to guarantee the ever greater flexibility and adaptability of this tool in the service of an ever increasing number of users.

Through their combined efforts, the European Commission and the Council of Europe would like to demonstrate the importance they attach, in the process of the European project, to dialogue and exchange of information and to the respect for and promotion of the cultural and educational diversity which is characteristic of Europe.

INTRODUCTION

Definition and aims

A thesaurus is a controlled vocabulary which is structured in hierarchical and associative relationships. It is designed for documentalists who, in storing and retrieving documents in a given information system, use key words to refer to the content of the documents.

The *European Education Thesaurus* (EET) constitutes an indexing language which was developed in accordance with ISO 5964 1985 on the 'Guiding principles for the establishment and development of multilingual thesauri'.

The EET is now available in all eleven official languages of the European Union. It has also been chosen as a documentary language in several countries of Central, Northern and Eastern Europe. Versions in Albanian, Croatian, Czech, Hungarian, Polish, Slovenian and Turkish have been developed in parallel.

The truly multilingual character of the EET is an essential part of its value. In accordance with international standards, all languages have equal status, that is to say, every concept is represented by a descriptor in each of the languages. The particular character and diversity of each of the eleven European Union languages are taken into account by means of scope notes and non-descriptors, which vary in content and number from one language to another.

Strict equality among languages is likewise respected in the updating of the content where, as necessary, the process is supported by transnational cooperation within the same language region.

This multilingual approach has made the EET the preferred means of communication for education specialists in a number of countries for exchanges of information or joint enlargement of the databases belonging to national and international documentation systems.

The EET has proved particularly useful for the indexing of educational research data contained in EUDISED (European Documentation System for Education in Europe), the database of the Council of Europe, which is accessible on the Internet via the Biblioteca di documentazione pedagogica in Florence (http://www.bdp.it/banche/banche.html, select: EUDI) and the databases of EURYDICE (the EU Network for information on education in Europe) and in particular its Community database on the education systems (EURYBASE – http://www.eurydice.org).

Moreover, all European countries consider the *European Education Thesaurus* to be the most appropriate indexing language for their documentation centres and information systems on education.

1998

Field

The *European Education Thesaurus* covers various subject areas which together constitute the interdisciplinary field of education:

– principles and systems of education,
– educational policy,
– educational institutions, teachers, pupils,
– school buildings and teaching equipment,
– curricula and teaching subjects,
– educational research, information on teaching,
– psychology of education,
– sociology of education,
– economics of education,
– public administration.

The EET also contains names of countries and certain regions.

The need to represent comparable real-life situations led to the introduction of concepts absent from a given educational system or language region. In such cases, the corresponding descriptor in the language version concerned was 'lent' to other language versions.

The purpose of this practice is to improve mutual knowledge of the different educational systems.

Similarly, in response to the needs of the EUDISED database of the Council of Europe which requires detailed vocabulary in the field of educational research, the EET has sought to respond to various specialist research requirements.

Background

The *European Education Thesaurus* is the outcome of 17 years of cooperation between the Council of Europe and the Commission of the European Communities.

In 1981, the General Secretaries of these two organisations signed a Cooperation Agreement for the ongoing development and joint financing of the EUDISED Thesaurus (later called *European Education Thesaurus*), the first edition of which had been prepared in 1974 under the aegis of the Council of Europe.

In 1984, cooperation between the two organisations produced the second edition of the EUDISED Thesaurus, consisting of one language version for each of the European Community's then nine official languages. For this edition, the vocabulary of the Thesaurus was completely restructured and new descriptors, collected between 1974 and 1979, were added.

Between 1986 and 1991, a substantial amount of updating took place in close cooperation between the Council of Europe and the EURYDICE European Unit – acting on behalf of the Task Force for Human Resources, Education, Training and Youth of the Commission of the European Communities.

In 1991, the third edition, now entitled the *European Education Thesaurus*, was published by the Office for Official Publications of the European Communities.

Since 1991, annual meetings of the EET-Management Group, comprising one representative per language area, took place under the chairmanship of Michael Larsen (1991-95) and Elfriede Tajalli (1995-). The EET-Management Group agreed to a number of amendments. In a series of

Council of Europe documents (DECS/Doc (94) 24, 25, etc.), certain amendments were published. At further meetings of the EET-Management Group (Vienna 1994, Prague 1995, Budapest 1996), new amendments were accepted.

The present fourth edition therefore incorporates all these addenda into the 1991 version and adds the two new official languages of the European Union to which those responsible for these language versions have made an active contribution.

Updating process for the fourth edition

The aim of this fourth edition has thus been to provide all the users of the *European Education Thesaurus* with a new complete version and to include the two new languages of the European Union (Finnish and Swedish). It has been financed by the European Commission through the European Unit of EURYDICE, which has coordinated the work in close cooperation with the Secretariat of the Council of Europe. The work has been undertaken by the expert, Mr Corentin Roulin of BJL-Information Technology, who acted as the Council of Europe's Thesaurus consultant between 1991 and 1996. Distribution of the EET on the Internet has been undertaken by the European Unit of EURYDICE.

While the previous editions had been published in hard copy, this fourth edition will be distributed largely through the Internet on EURYDICE's web site (http://www.eurydice.org). As this documentary language is essential to the mutual understanding between educational systems, the aim is to make it accessible for the greatest possible number of interested users. These will be able to express their views and make suggestions through a reply coupon provided on the site.

A great many changes were made once the updating of the content was completed: 203 new descriptors were added, 10 were removed (see annexes), the hierarchical order was changed which meant altering the corresponding microthesauri, scope notes or non-descriptors were added or removed depending on the language version, and corrections were made in the wording of terms in all language versions.

1998

DESCRIPTION

Language versions

The *European Education Thesaurus* comprises eleven language versions, each language being identified by a language code:

Language	Language code
Danish	DA
German	DE
Greek	EL
English	EN
Spanish	ES
Finnish	FI
French	FR
Italian	IT
Dutch	NL
Portuguese	PT
Swedish	SV

In accordance with international standards:

–   all languages in the thesaurus have equal status, that is to say each descriptor in any language must correspond to a descriptor in each of the other languages;

–   there is not necessarily any one-to-one equivalence between non-descriptors; indeed this is not possible since the lexicological wealth of a language varies from subject to subject.

Structure of the Thesaurus

In accordance with established standards, the *European Education Thesaurus* contains:

–   **Descriptors,** i.e. terms which must be used to represent concepts when indexing documents and formulating questions.

   Example 1:

   pupil

–   **Non-descriptors,** i.e. synonyms or terms designating concepts very close to those represented by the descriptors and providing points of access to the EET, whereby the terminology used in indexing and formulating questions is directed towards the descriptors as the preferred terms; the number of non-descriptors may vary from one language to another and they always appear in italics.

   Example 2:

   school child

–   **Scope notes,** briefly explain the use for which the descriptor is intended where it is ambiguous; scope notes may vary in number and content from one language to another; they are introduced by a symbol common to all thesaurus languages – SN (= SCOPE NOTE) – which is placed between the descriptor and its scope note.

1998

Example 3:

Pupil
SN   person receiving primary, lower and upper secondary education

– **Semantic equivalence relations** between descriptors and non-descriptors are shown by symbols which are identical for all thesaurus languages:

    – USE (= use), placed between the non-descriptor and the corresponding descriptor.

        Example 4:

        school child
        USE   pupil

        N.B.: a non-descriptor must refer to a single descriptor.

    – UF (= USED FOR), placed between a descriptor and the non-descriptor(s) which it represents.

        Example 5:

        pupil
        UF   school child

        N.B.: any descriptor may have zero, one, two or more non-descriptors corresponding to it.

– **Hierarchical relations** between descriptors, shown by symbols:

    – BT (= BROADER TERM), placed between a specific descriptor and a generic descriptor.

Example 6:

teaching language
  BT1 language policy

– NT (= NARROWER TERM), placed between a generic descriptor and a specific descriptor.

Example 7:

pupil
  NT1 boarder

– The various hierarchical levels are shown by a digit to the right of the symbol BT or NT, indicating the number of hierarchical levels separating the terms in question.

Example 8:

biology

BT1 biological science
      BT2 natural sciences
NT1 physiology
      NT2 neurophysiology
      NT2 psychophysiology
       NT3 psychosomatics

– **Associative relationships** of ideas between descriptors, shown by the symbol RT (= RELATED TERM = associated term or *see also),* placed between two associated descriptors.

Example 9:

pupil
RT parent-pupil relation

Apart from such standardised semantic relations, the EET includes:

– **Relations of inclusion** in a microthesaurus: each microthesaurus is represented by a two-figure serial number followed by the title of the microthesaurus. This number is identical in every language.

N.B.: The name of the microthesaurus accompanied by its serial number does not constitute a descriptor.

A relation of inclusion is shown by the symbol MT (= MICROTHESAURUS) followed by the serial number and title of the microthesaurus, this indication being repeated as often as necessary.

As a multiple hierarchy is permissible, any descriptor may appear in more than one microthesaurus.

Example 10:

school success
MT 06   evaluation
MT 23   behaviour incentive

– **Relations of linguistic equivalence** between descriptors designating the same concept in different languages is indicated below each descriptor in a given language by the corresponding descriptors in the other languages, each of which is preceded by a language indicator.

Example 11:

ability
da: evne
de: Fähigkeit
el: ?a??t?ta
es: capacidad
fi: kyvykkyys
fr: capacité
it: capacità
nl: capaciteit
pt: Capacidade
sv: förmåga

As there is no correspondence between non-descriptors from one language to another, relations of linguistic equivalence appear under descriptors only.

The entry year is indicated for descriptors as of the 1991 version.

Characteristics

All the language versions of the *European Education Thesaurus* comprise:

– 42 microthesauri
– 2953 descriptors

Each language version comprises a variable number of scope notes and non-descriptors.

Language version	Number of	
	Scope notes	Non-descriptors
DA	236	380
DE	262	465
EL	270	479
EN	293	558
ES	292	520
FI	131	20
FR	322	616
IT	347	587
NL	215	1006
PT	304	368
SV	2	1

Methods of presentation

The hard copy version of the *European Education Thesaurus* is published in one volume for each language version. Each volume has four parts, in addition to the preface and introduction:

– a structured alphabetical presentation;
– grouping by semantic fields (microthesauri);
– a rotated alphabetical presentation;

– an annex indicating modified descriptors and descriptors deleted from the current edition.

In the alphabetical presentation, the page number is accompanied by the wording of the term (descriptor or non-descriptor):

– uneven page numbers are preceded by the wording of the last term appearing under the heading 'entry' on that page,
– even page numbers are followed by the wording of the first term appearing under the heading 'entry' on that page.

Structured alphabetical presentation

Structured alphabetical presentation, with three columns per page, includes two types of 'entries' or 'information blocks':

– descriptor
– non-descriptor

listed alphabetically according to the complete designation of the descriptors and non-descriptors.

* **Descriptor entry**

– wording of the descriptor, in bold letters;
– year of introduction (since 1991) in brackets;
– number(s) and designation(s) of the microthesaurus or microthesauri to which the descriptor refers, in italics and preceded by the symbol MT (MICROTHESAURUS);
– language equivalents in each of the ten other languages; each equivalent is preceded by the relevant language code; the language equivalents are presented in the alphabetical order of the language codes, i.e. in the same order in all language versions;
– scope note, where the descriptor has one, preceded by the symbol SN (SCOPE NOTE);
– non-descriptor(s) relevant to the entry descriptor, in italics and preceded by the symbol UF (USE FOR);

– generic descriptor(s) of the entry descriptor, preceded by the symbol BT (BROADER TERM) and a figure showing the number of hierarchical levels between the generic descriptor and the entry descriptor; generic descriptors are listed:
  – firstly, in order of ascending hierarchical levels;
  – secondly, within each hierarchical level in alphabetical order;
– specific descriptor(s) of the entry descriptor preceded by the symbol NT (NARROWER TERM) and a figure showing the number of hierarchical levels; specific descriptors are listed:
  – firstly, in order of descending hierarchical levels;
  – secondly, within each hierarchical level in alphabetical order;
– descriptor(s) associated with the entry descriptor preceded by the symbol RT (RELATED TERM) and listed alphabetically.

Example 12:

vocational guidance

> *MT 33 labour environment*
> da: erhvervsvejledning
> de: Berufsberatung
> el: επαγγελματικός προσανατολισμός
> es: orientación profesional
> fi: ammatinvalinnanohjaus
> fr: orientation professionnelle
> it: orientamento professionale
> nl: beroepskeuzebegeleiding
> pt: Orientação profissional
> sv: yrkesvägledning
> SN   help given inside or outside the school system in choosing a career or occupation
UF   *occupational guidance*

BT1   transition from school to work
NT1   career profile
  NT2   job classification
  NT2   job description
  NT2   occupational choice
  NT2   occupational prognosis
RT   guidance

*   Non-descriptor entry

– wording of the non-descriptor, in bold italics;
– corresponding descriptor, preceded by USE.

Example 13:

school child
USE pupil

Systematic presentation – Microthesaurus

The systematic presentation, with three columns to the page, provides all the 'hierarchical links' contained in the microthesaurus in numerical order. This order is identical from one language to the other.

The microthesaurus titles, preceded by their number in brackets, are given in bold type in a frame at the head of the first or second column of a page (where the previous microthesaurus ends on the first column of a page); this title also appears in the page heading. No page has more than one microthesaurus title.

Within a microthesaurus, the 'hierarchical links' are presented in the alphabetical order of the 'top terms'. This order therefore varies from one language to the other.

The top term is in bold and its specific descriptors are presented below it, following the same principles of classification as in the presentation of the specific descriptors in the structured alphabetical presentation, i.e.:

–   first, in descending hierarchical order;
–   within each hierarchical level, in alphabetical order.

The hierarchical relationships are indicated by offsetting the specific descriptors to the right by as many points as there are hierarchical levels separating them from their 'top terms'.

Every descriptor ('top term' or specific descriptor) is also accompanied by, as necessary, the following information:

–   on the same line: an indication of multiple hierarchical relationships, shown by a double arrow where the descriptor has another origin under another microthesaurus; this is followed by the number of the other microthesaurus in brackets;
–   on one or more additional lines, proportionately offset to the right: associated descriptor(s) preceded by the code RT (Related Term) and set out in alphabetical order. If an associated descriptor also belongs to one or more other microthesauri, the relevant number(s) appear(s) in brackets after the definition.

Where a hierarchical sequence goes beyond a column, the break is indicated by an arrow pointing to the right at the bottom of the column and an arrow pointing to the left at the top of the next column on the same or the following page.

Non-descriptors do not appear here.

Rotated alphabetical presentation

The rotated alphabetical presentation, with two columns per page, comprises two types of 'entries' or 'information blocks':

–   descriptor,
–   non-descriptor,
–

alphabetically selected from each significant word (printed in bold) in the wording of the descriptors and non-descriptors.

If a term is artificially separated in order to isolate one of its words for insertion in an alphabetical list, this separation is indicated with a hyphen.

    Example 14:

intelligens- **kvotient** (20)   [Danish]
Musik- **erziehung** (10)   [German]
onderwijs- **experiment** (16)   [Dutch]

The information given under each of these entries is as follows:

*   **Descriptor entry**

–   indication of the descriptor,
–   number(s) of the microthesaurus or microthesauri containing the descriptor, indicated in brackets.

*   **Non-descriptor entry**

–   indication of the non-descriptor,
–   corresponding descriptor preceded by USE and followed by the number(s) of the microthesaurus or microthesauri containing it, indicated in brackets.

Example 15:

	child (31)
Abandoned	**child** (23, 31)
Abnormal	***child***
	USE exceptional (23)
Adopted	**child** (31)
Backward	**child** (26, 31)
Battered	**child** (31)
Illegitimate	**child** (26, 31)
Neglected	**child** (26, 31)
Only	**child** (31)
Problem	**child** (31)
Retarded	***child***
	USE backward child (26, 31)
Rights of the	**child** (36)
School	***child***
	USE pupil (04)

Pre-school	**child** (31)
Migrant worker's	**child** (31)
	child of alien laborer
	USE child of foreign national (31)
	child care (31)
	child care centre
	USE day nursery (07, 36)
	child development (19)

ANNEX: newly added descriptors and eliminated descriptors

The annex contains:
– the list of new descriptors,
– the list of descriptors which were deleted and which were not included as non-descriptors.

UTILIZATION

Instructions for the indexing of documents

Indexing is the operation whereby the concepts dealt with in a document are identified and represented by descriptors in the Thesaurus.

Its purpose is to make it possible to extract from all the bibliographical material the information which will give answers to the questions put; such extraction is possible only if the concepts in the question are formulated in the same language, i.e. that of the EET descriptors.

The selection of concepts to be represented is based on two rules:

–   **selectivity**: only those concepts must be selected for which the document supplies information likely to interest users;
–   **exhaustiveness**: all useful concepts contained in the text, whether explicit or implicit, must be selected.

Representation of the concepts thus selected, by means of EET descriptors, entails the application of two other rules:

–   **vertical specificity**: the descriptor must be on the same level of specificity as the concept or, failing this, on the level immediately above it in the EET;
–   **horizontal specificity**: a concept consisting of two or more words must be expressed by means of a composite descriptor, if there is one, rather than by a combination of single descriptors.

In concrete terms, representation of concepts by descriptors proceeds as follows:

–   if the document is in a foreign language, the concepts are translated into the language version of the EET used by the indexer;
–   'entries' corresponding to the concepts expressed in the document are looked for in the rotated alphabetical presentation of the EET.

There are three possibilities:

–   expression of the concept corresponds to a descriptor signifying it: this descriptor is then selected;
–   expression of the concept corresponds to a non-descriptor signifying it: the latter refers to the descriptor to be used; the above procedure then applies;
–   there is no entry corresponding to the expression of the concept. Two methods may then be used:

*   either think of another formulation of the concept in natural language and see whether it corresponds to an entry in the Thesaurus,

*   or define the general class or classes (microthesaurus) covering the recalcitrant concept, consult the microthesauri and look among available descriptors for those which best represent the concept.

Correct application of the rules of selectivity, exhaustiveness and specificity is essential to the quality of a documentary information storage and retrieval system.

1998

– **Selectivity** of indexing:

  – is independent of the quality of the EET,
  – depends on the indexing policy applied for each documentary system,
  – depends on the indexing documentalist's knowledge of the users and their needs.

– **Exhaustiveness** is conditioned:

  – by the wealth of content of the EET; obviously only those concepts which are present in the documentary language can be indexed;
  – by the indexing policy, which establishes the depth of indexing: this depth cannot be determined in advance; there is an optimum which depends on the nature of the questions and which can only be determined iteratively;
  – by the behaviour of the documentalists (in the absence of a specific indexing policy) and their meticulousness.

An increase in the indexing exhaustiveness influences information retrieval in two ways:

– it improves response, i.e. the percentage of relevant documents extracted from a given database in responding to a question;
– it lowers precision, i.e. the proportion of truly relevant documents extracted in response to a question.

If all concepts contained in the documents, even very subsidiary ones, are indexed, the chances are increased of retrieving all or almost all of the relevant documents in response to the questions put, but at the same time the bibliographies obtained will contain a high percentage of documents which are of little or no relevance.

– **Specificity** of indexing also depends on:

  – the EET, which may or may not include a large number of specific descriptors;
  – the indexing policy;
  – the documentalists.

Unlike increased exhaustiveness, an increase in specificity results in:

– fewer documents retrieved,
– greater precision.

If documents are indexed by means of descriptors having the same level of specificity as the concepts which they contain, questions concerning these descriptors will produce a bibliography that will not include those documents which are indexed at a more generic level, some of which in fact will contain relevant information. By way of compensation, however, the list will include few irrelevant documents. It should be noted, however, that documents indexed at the more generic level may be retrieved precisely by searching among these descriptors.

The reverse is unfortunately not the case. If indexing is not specific, the result is:

– more documents retrieved,
– lower precision.

For that reason, it is important to be specific when indexing documents.

Instructions for the formulation of questions

The formulation of a question and indexing are symmetrical operations: indexing consists of identifying the concepts dealt with in a document, on which information likely to be of interest to users is given; these concepts are then designated by descriptors. Formulation of a question consists of identifying the descriptors which may have served to translate the concepts that represent information responding to the question, and subsequently extracting the indexed documents by means of the descriptors.

Identification of the concepts in a question calls for a very different approach from that used in identifying the concepts in a document: the indexing of a document is an operation that leads to a high degree of compression. A document of 5 to 10 pages is generally represented by 5 to 15 descriptors (i.e. a compression rate of 1% and 1 per mil), while the formulation of a question usually leads to an expansion of the initial wording.

The first stage in the formulation of a question is to indicate as explicitly as possible the information requirement, expressed in natural language. As a second step, the concepts which together comprise the wording of the question are identified.

Then these concepts have to be represented by means of the EET descriptors. These cases should be distinguished:

–   The question is precise and the concepts it contains are present in the EET; one will preferably make use of the complete structured alphabetical list in which one will find:
    –   the descriptors or non-descriptors (and, from there, other descriptors) corresponding to the wording of the concepts in the question;
    –   descriptors in hierarchical relations and associative relationships which may help the user to broaden the question;
–   none or not all of the concepts contained in the question appear in the EET: after failing to find the concepts in the structured alphabetical list or in the rotated alphabetical list, the documentalist will explore the microthesauri corresponding to the field(s) covered by the question and search for existing descriptors designating concepts nearest to those expressed in the question;
–   the question is imprecise, as is frequently the case: the documentalist will preferably use microthesauri, exploring the one(s) which correspond(s) to the field(s) of interest; there one will find an ordered collection of descriptors covering the field(s) concerned and will be able to choose the most suitable descriptors.