



# Guidelines for Multilingual Thesauri

Working Group on Guidelines for Multilingual Thesauri  
Classification and Indexing Section, IFLA  
April 2005

Chair: Gerhard J.A. Riesthuis (Netherlands)  
Members: Lois Mai Chan (USA), Patrice Landry (Switzerland),  
Pia Leth (Sweden), Ia McIlwaine (United Kingdom), Martin Kunz (Germany),  
Dorothy McGarry (USA), Max Naudi (France), Marcia Lei Zeng (USA)

## Table of Contents

|   |    |
|---|----|
| 1. Introduction .....   | 1  |
| 2. List of abbreviations .....                                | 2  |
| 3. Building multilingual thesauri starting from scratch ..... | 2  |
| 3.1 Introduction  |    |
| 3.2 Structure   |    |
| 3.3 Morphology and Semantics                                  |    |
| 4. Starting from existing thesauri .....                      | 13 |
| 4.1 Merging   |    |
| 4.2 Linking   |    |
| 5. Glossary .....   | 15 |
| 6. Literature .....   | 18 |
| Appendix A. Example of a non-symmetrical thesaurus .....      | 20 |

# 1 Introduction

Multilingual information retrieval languages exist in different forms, e.g. subject headings lists, thesauri, enumerated classifications, analytico-synthetic classifications and so on.

In a multilingual information retrieval language both the terms and the relationships are represented in more than one language.

In this document the emphasis is on multilingual thesauri.

The objective of this document is to add to the existing Guidelines for Multilingual Thesauri as worded in the ISO-standard for multi-lingual thesauri (ISO-5964-1985) or in handbooks on thesaurus building, such as Aitchison (2000). The general principles for the building of monolingual thesauri are assumed.

Since the drawing up of the Guidelines in the 1970s two developments have played important roles in the thinking about multilingual access to information: the building of non-symmetrical thesauri (see § 2.2) and the linking of two or more thesauri and/or controlled vocabularies (see § 3).

There are three approaches in the development of multilingual thesauri:

1. building a new thesaurus from the bottom up
  - a. starting with one language and adding another language or languages
  - b. starting with more than one language simultaneously
2. combining existing thesauri
  - a. merging two or more existing thesauri into one new (multilingual) information retrieval language to be used in indexing and retrieval
  - b. linking existing thesauri and subject heading languages to each other; using the existing thesauri and/or subject heading languages both in indexing and retrieval.
3. translating a thesaurus into one or more other languages.

In the last case the languages involved are not treated equally. The language of the existing thesaurus becomes the dominant language<sup>1</sup>. This approach is not treated in this document.

Linking is typically used in situations where different agencies are using their own indexing vocabularies in their own languages for their own information systems. The linking makes it possible for the end-user to search in all linked information retrieval systems using any one of the linked information retrieval languages. An example of a multilingual linking project is the MACS project (see <http://infolab.kub.nl/prj/macs>).

Building from the bottom up is only viable in cases where a new information language is envisaged. The main advantage is that the languages involved can be treated equally.

---

<sup>1</sup> See about translation of thesauri: Nase & Mdivani (1996).

In both approaches dealt with in this document two groups of problems are encountered:

1. Semantic problems

Semantic problems pertain to equivalence relations between terms used as preferred and non-preferred terms in information retrieval languages. Equivalence relations exist not only within each separate language involved, but also between the languages (intra-language equivalence and inter-language equivalence). Intra-language homonymy and inter-language homonymy are also considered semantic questions. Additional problems pertaining to semantics involve the scope, form and choice of thesaurus terms.

2. Structural problems

Structural problems involve hierarchical and associative relations between the terms. An important question in this respect is whether the structure should be the same or different for each language. In most if not all cases of linking, the structure will most probably not be the same in all the information retrieval languages involved. In the other approaches mentioned it is possible in principle to apply the same structure to all languages. This question will be discussed later.

A glossary appears at the end of this document.

## 2 List of abbreviations<sup>2</sup>

| <i>Dutch</i> | <i>English</i> | <i>German</i> | <i>French</i> | <i>Meaning</i>       |
|--------------|----------------|---------------|---------------|----------------------|
| USE          | USE            | BS            | EM            | Use term ... instead |
| UF           | UF             | BF            | EP            | Used for ...         |
| BT           | BT             | OB            | TG            | Broader term         |
| NT           | NT             | UB            | TS            | Narrower term        |
| RT           | RT             | VB            | TA            | Related term         |
| SN           | SN             | D             | NE            | Scope note           |

## 3 Building multilingual thesauri starting from scratch

### 3.1 Introduction

The morphology of the descriptors and non-descriptors has been discussed at great length in guidelines for monolingual thesauri<sup>3</sup> in *Principles Underlying Subject Heading Languages (SHLs)*<sup>4</sup> and in the context of the MACS project<sup>5</sup>

<sup>2</sup> Meaning of the German abbreviations is BS: Benutze; BF: Benutzt für; OB: Oberbegriff; UB: Unterbegriff; VB: Verwandter Begriff; D: Definition. The meaning of the French abbreviations is : EM: Employer; EP: Employer pour; TG: Terme générique; TS: Terme spécifique; TA: Terme associé; NE: Note explicative. Instead of EM also *Voir* is used, instead of NE one finds also NA: Note d'application. In Dutch the English abbreviations are used.

<sup>3</sup> For an overview of such guidelines see Milstead (2001).

In this document only a few remarks about morphological problems will be made. Greater attention will be given to equivalence relationships, with emphasis on inter-language equivalence.

Structural problems form a major part of the problems discussed in this document.

### 3.2 Structure

Two approaches to the semantic structure of multilingual thesauri can be distinguished. The most common view is that all different language versions of a multilingual thesaurus have to be identical and symmetrical; each descriptor must have one and only one equivalent in every language and be related in the same way to other descriptors in the given language (a *symmetrical thesaurus*).

The alternative is a non-identical and non-symmetrical structure where the number of descriptors in each language is not necessarily the same and also where the way descriptors are related to each other can be different for the different languages (a *non-symmetrical thesaurus*).

Builders of a *symmetrical thesaurus* aim at full correspondence between descriptors and relations. This means that each descriptor in any of the languages has an equivalent in all other languages and that the relations between the descriptors in all languages are the same. If in language X a generic relation exists between descriptor A and B, then a generic relation between the equivalents A` and B` also exists in language Y. As a consequence it can happen, and mostly does happen, that cross-language equivalences are forced where they do not exist and questionable relational structures occur.

An example of the complexity of the network of relations is given in Hudon (2001):

| <i>English (source)</i> | <i>French (target)</i>                 |
|-------------------------|--|
| EDUCATION               | ÉDUCATION<br>ENSEIGNEMENT<br>FORMATION |
| TEACHING                | ENSEIGNEMENT                           |
| INSTRUCTION             | ENSEIGNEMENT<br>INSTRUCTION            |

The meaning of the English term EDUCATION is broader than that of the French term ÉDUCATION or, expressed differently, EDUCATION and TEACHING are less clearly distinct in English than ÉDUCATION and ENSEIGNEMENT are in French. This fact has consequences for the generic relations between these terms and the hierarchical narrower terms under each of the descriptors (see Appendix A).

In symmetrical multilingual thesauri the meaning of terms often has to be made broader or narrower than in the natural use of that language but, even so, artificial or “coined” terms cannot be avoided.

---

<sup>4</sup> Principles (1999)

<sup>5</sup> Landry (2004)

### 3.3 Morphology and Semantics

#### 3.3.1 Scope of descriptors

Each descriptor included in a thesaurus should represent a single concept<sup>6</sup>. It can be a single-word term or a multi-word compound term. In principle, the descriptors should be terms from a natural language, and no artificial terms should be adopted for the thesaurus. The last point is discussed in more detail below.

The concepts represented by descriptors can be grouped into mutually exclusive categories known as facets, based on shared characteristics. Entities, Processes, Attributes, Agents, Geographical places and Chronology are often used as facets.

Unique entities or “classes-of-one” concepts are expressed as proper nouns. Often official or unofficial versions in different languages exist. A choice has to be made between

- choosing the proper name in one of the languages to be used for all languages
- using the proper name in each language that has the name, and the proper name in the original language for the languages that lack it
- using translations as far as is sensible and possible.

In all cases appropriate non-preferred terms should be added. It is advisable to use an appropriate name authority file or list for each language.

| <i>English</i>                                    | <i>French</i>                                | <i>Dutch</i>   |
|---|--|--|
| International Committee of the Blue Shield (ICBS) | Comité International du Bouclier Bleu (CIBB) | Internationaal Comité van het Blauwe Schild<br>UF International Committee of the Blue Shield (ICBS)<br>UF Comité International du Bouclier Bleu (CIBB) |

The English and French descriptors are in official forms of the name of this “class-of-one”; in Dutch the descriptor is a translation with the official English and French terms as non-descriptors<sup>7</sup>.

#### 3.3.2 Clarification and disambiguation of descriptors

The scope of each descriptor is limited to one meaning within the domain of the thesaurus. This meaning is not necessarily the most common meaning associated with the term in the natural language. Clarification of the meaning may be provided by the context through the structure in which the descriptor is placed. Also, the descriptor itself should be formulated in such a way that it conveys the intended scope in any language to any user who is familiar with the language in question. Ambiguity should be avoided as far as possible. A qualifier (see below) is often sufficient for disambiguation purposes.

---

<sup>6</sup> See about concept Dahlberg, 1989.

<sup>7</sup> See <http://www.ifla.org/blueshield.htm>

### 3.3.3 Homographs and qualifiers

Homographs occur not only within one language, but also between languages. Preferably one or more of the homographs should be replaced by a commonly used synonym if available. Scope notes should be added when appropriate.

When homographs are needed as thesaurus terms, the meaning of each should be clarified by adding a qualifier in parentheses. An example of the first case is *CRANES (BIRDS)* and *CRANES (LIFTING EQUIPMENT)*. The fact that *CRANES* is a homograph in English does not necessarily mean that equivalent terms in other languages are also homographs. The Dutch term *KRANEN* is a homograph too, but with the meanings *CRANES (LIFTING EQUIPMENT)* and *TAPS*.

| <i>English (British)</i>   | <i>English (USA)</i>       | <i>Dutch</i>  | <i>French</i>                                   |
|----------------------------|----------------------------|---|---|
| cranes (birds)             | cranes (birds)             | kraanvogels   | grue (oiseau)                                   |
| cranes (lifting equipment) | cranes (lifting equipment) | hijskranen<br>SN voor andere typen kranen, zie aldaar           | grue (appareil de levage)                       |
| water taps                 | water faucets              | waterkranen   | robinet à eau                                   |
| gas taps                   | gas faucets                | gaskranen   | robinet à gaz                                   |
| taps                       | faucets                    | kranen<br>SN voor kranen als hijswerktuig<br>gebruik hijskranen | robinet<br>NT robinet à eau<br>NT robinet à gaz |
| NT water taps              | NT water faucets           | NT waterkranen  |   |
| NT gas taps                | NT gas faucets             | NT gaskranen  |   |

Inter-language homographs will not normally lead to confusion in a multilingual thesaurus, but exceptions can exist, as in the next example. Scope notes can be desirable or necessary.

| <i>Dutch</i>  | <i>English</i>              | <i>German</i>                                     | <i>French</i>                  |
|---|-----------------------------|---|--------------------------------|
| GESCHENKEN<br>UF <i>GIFTEN</i><br>UF <i>CADEAUS</i><br>UF <i>KADO'S</i> | GIFTS<br>UF <i>PRESENTS</i> | GESCHENKE<br>NT <i>GABEN</i><br>NT <i>SPENDEN</i> | CADEAUX<br>UF <i>DONATIONS</i> |
| VERGIFTEN   | POISONS                     | GIFTE   | POISONS                        |

An exception is the case where terms in more than one language are given in one alphabetical order. Add the language in parentheses in such cases.

An example from LCSH (English) and RAMEAU (French):

(eng) CHANCE = HASARD (fre)

(eng) FORTUNE = CHANCE (fre)

In the alphabetical index we get:

...  
 CHANCE (eng)  
 CHANCE (fre)  
 FORTUNE  
 HASARD  
 ...

### 3.3.4 Forms of terms

The forms of the terms in each language should be based on the conventions in the languages involved. A descriptor should preferably consist of a noun or noun phrase. Verbal nouns are acceptable. Use the form customary in the languages involved such as gerunds in some languages like English, or infinitives in some other languages like Dutch and Romanian.

#### **Example**

| <i>English</i> | <i>Dutch</i> | <i>Romanian</i> |
|----------------|--------------|-----------------|
| Swimming       | zwemmen      | înota           |

Noun phrases are compound terms, and occur in two forms:

a) adjectival phrases

| <i>English</i>     | <i>Dutch</i>        |
|--------------------|---------------------|
| concrete bridges   | betonnen bruggen    |
| library catalogues | bibliotheekcatalogi |

Note that the phrase LIBRARY CATALOGUES becomes one (compound) word in Dutch!

b) prepositional phrases

| <i>English</i>                      | <i>Dutch</i>                    |
|-------------------------------------|---------------------------------|
| accessories after the fact          | medeplichtigen                  |
| hospitals for children <sup>8</sup> | kinderziekenhuizen              |
| prisoners of war                    | krijgsgevangenen                |
| very large scale integration        | integratie op zeer grote schaal |

Note also here that some English phrases are equivalent to Dutch compound words.

The Dutch term “medeplichtigen” means accessories (accomplices) in general. There is no direct equivalence for “accessories after the fact”. In a symmetric thesaurus the solution is:

| <i>English</i>                                      | <i>Dutch</i>   |
|---|----------------|
| Accessories<br><b>UF</b> Accessories after the fact | Medeplichtigen |

<sup>8</sup> This phrase is used here for the sake of the example. CHILDREN’S HOSPITAL is more common. Also the Dutch language has ZIEKENHUIZEN VOOR KINDEREN as a less used equivalent.

In a non-symmetric thesaurus the solution is:

| <i>English</i>                       | <i>Dutch</i>   |
|--------------------------------------|----------------|
| Accessories                          | Medeplichtigen |
| <b>NT</b> Accessories after the fact |                |

Prepositions can add unnecessary length and clumsiness to a term. In most languages they should therefore be avoided if possible. For example, use CARBOHYDRATE METABOLISM rather than METABOLISM OF CARBOHYDRATES.

Regarding the use of the plural or singular of nouns, follow the conventions of the different languages (recommended in cases where national standards exist) or choose to use the same rules for number in all languages.

Avoid the use of adjectives (without a noun), verbs and initial articles. If necessary, follow the appropriate conventions of the individual languages.

Give all terms in the script of the individual languages. Use the appropriate capitalization, punctuation, diacritics and special characters of the individual languages. For spelling, the official rules of the languages should be followed. For alphabetization of terms the rules of the different languages should be used.

The conventions used for the forms of the terms should be clearly explained in the introduction to the thesaurus.

### 3.3.5 Compound terms

A compound term is a term consisting of more than one word (a phrase term), or a compound word. Compound words are words composed of more than one word (often nouns), with or without one or more letters connecting the words that compose the compound word, e.g. the first s in the German compound word ‘Zukunft’sangst’ that consists of ‘Zukunft’ und ‘Angst’ and means ‘fear of the future’<sup>9</sup>.

Typically, a compound term pre-coordinates two or more simple concepts into one combined concept. The use of compound terms in a thesaurus tends to increase its level of specificity. Specificity increases the precision with which a particular concept can be accurately represented and consequently increases the facility with which unwanted documents can be excluded.

The parts of most compound terms can be distinguished as follows:

a) the **focus** or **head**, i.e. the noun component that identifies the general class of concepts to which the term as a whole refers.

Examples:

- 1) the noun component “indexes” in the phrase term “printed indexes”.
- 2) the noun “hospitals” in the prepositional phrase “hospitals for children”.
- 3) the part “band” in the compound word “broadband”.

b) The **difference** or **modifier**, i.e. one or more further components that serve to narrow the extension of the focus and so specify one of its subclasses.

Examples:

---

<sup>9</sup> In English different forms of the same compound terms can exist next to each other: separate words, words connected with hyphens or words written as a compound term.



- 1) the adjective “printed” in the compound term “printed indexes”.
- 2) the preposition-plus-noun combination “for children” in the compound term “hospitals for children”
- 3) the part “boek” in the compound word “Boekband”.

The question of whether to admit a compound term is dealt with in detail in guidelines for monolingual thesauri.

In general, if a compound term is split in one language, equivalents in other languages should be split too. If compound terms – especially compound words – are split, and common to one or more languages, a reference by means of a non-descriptor should be made. An example in the Dutch language:

AUTOMOTORS  
USE AUTOS + MOTORS

The form of a compound term is not always the same in different languages; for example, the English term RAIL SAFETY is VEILIGHEID BIJ SPOORWEGEN in Dutch, a prepositional term not very adequate as a descriptor. In a non-symmetric thesaurus, a possibility would be:

| <i>English</i> | <i>Dutch</i>   |
|----------------|--|
| RAIL SAFETY    | veiligheid bij spoorwegen<br>USE VEILIGHEID + SPOORWEGEN |

This possibility has consequences for the structure of the thesaurus (see § 2.2).

### 3.3.6 Equivalence

Inter-language equivalence has three aspects: semantic, cultural and structural. The semantic and cultural aspects refer to the meaning of the terms and the way the terms are used in a given language or culture. The structural aspect relates to the hierarchical and associative relations.

Semantic and cultural equivalence is only relevant for descriptors; in most multilingual retrieval languages no attempts are made to give equivalent terms for all non-descriptors. This practice is recommended to avoid artificial terms in one or more languages. Several different cases of semantic and cultural equivalence between preferred terms in the languages X and Y can be distinguished:

1. *Exact equivalence (interlinguistic synonymy)*<sup>10</sup>: the descriptors in X and Y are semantic and culturally equivalent.

| <i>German</i>    | <i>English</i>     | <i>Dutch</i> |
|------------------|--------------------|--------------|
| Schienennetz     | Railroad network   | spoorwegnet  |
| UF Eisenbahnnetz | UF railway network |              |

2. *Inexact or near equivalence (inter-language quasi-synonymy, with a difference in viewpoint)*: the descriptors in X and Y express the same general concept but the meanings of the terms in X and Y are not exactly identical. Often the differences are more cultural

---

<sup>10</sup> The terms in brackets are taken from Hudon (1997). Several examples are taken from existing thesauri. This does not mean approval or disapproval of these thesauri.

than semantic, i.e. there is a difference in connotation or appreciation.<sup>11</sup> In the case of inexact equivalence the terms can be treated as if they were exact equivalents.

*Example*<sup>12</sup>

| <i>English</i>       | <i>Spanish</i>             | <i>French</i>        |
|----------------------|----------------------------|----------------------|
| historic settlements | = asentamientos historicos | ≈ site de peuplement |

3. *Partial equivalence (interlinguistic quasi-synonymy, with a difference in specificity)*: the descriptor in one of the languages has a slightly broader or narrower meaning than the descriptor in the other language.

There are three possible solutions:

- Treat the terms as exact equivalents.

*Example*<sup>13</sup> :

| <i>German</i> | <i>English</i> |
|---------------|----------------|
| Wissenschaft  | Science        |

- Adopt the terms from each language as loan terms in the other languages and organize these terms hierarchically; e.g., one term is designated as the broader term and one as the narrower term.

*Example*<sup>14</sup> :

| <i>German</i>  | <i>English</i>   |
|--|--|
| WISSENSCHAFT   | WISSENSCHAFT<br>SN Loan term adopted from German<br>NT SCIENCE |
| SCIENCE<br>D Lehnwort aus dem Englishen<br>OB WISSENSCHAFT | SCIENCE<br>BT WISSENSCHAFT                                     |

- Treat the situation as *single-to-many equivalence*.

4. *Single-to-many equivalence (too many or not enough terms)*: to express the meaning of the descriptor in one of the languages, two or more descriptors are needed in the other language.

Two different cases exist:

- the target language includes more than one equivalent of the source term (too many target terms).

<sup>11</sup> Strong-willed and pig-headed both means stubborn, but the first term has a positive connotation, the second a negative one.

<sup>12</sup> This example is taken from the introduction to the HEREIN thesaurus [www.European-heritage.net/sdx/herein/thesaurus/introduction.xsp] [June 2003]

<sup>13</sup> The example is taken from Schott (2002)

<sup>14</sup> From ISO 5964-1985, p. 12.

There are three possible solutions:

- a. create a precombined descriptor in the target language.

*Example*<sup>15</sup>

| <i>English</i>  | <i>French</i>   |
|-----------------|---|
| LISTED BUILDING | EDIFICE INSCRIT + EDIFICE CLASSE <sup>16</sup><br>EP EDIFICE INSCRIT<br>EP EDIFICE CLASSE |

- b. modify or specify the source term, e.g. by addition of a qualifier.

*Example*<sup>17</sup>:

| <i>English</i>                    | <i>German</i> |
|-----------------------------------|---------------|
| FESTIVAL (FEAST)                  | FEST          |
| RT HOLIDAY                        | RT FEIERTAG   |
| FESTIVAL (SERIES OF PERFORMANCES) | FESTIVAL      |

- c. establish one or more non-descriptor(s) in the target language, with links to the descriptor.

*Example:*

| <i>English</i> | <i>German</i>                     |
|----------------|-----------------------------------|
| FESTIVAL       | FESTIVAL<br>BF FEST<br>VB HOLIDAY |
|                | FEST<br>BS FESTIVAL               |

Solution c is the most acceptable solution from the point of view of language equality because no artificial term is created.

- The target language can only represent the source concept through a combination of terms (not enough target terms).

- a. make a precombined descriptor in the target language.

*Example:*<sup>18</sup>

| <i>German</i>   | <i>English</i>                               |
|-----------------|--|
| BERUFSPRAKTIKUM | VOCATIONAL EDUCATION +<br>PRACTICAL TRAINING |

<sup>15</sup> Example taken from the HEREIN thesaurus (see note 12).

<sup>16</sup> Note that “ÉDIFICE INSCRIT + ÉDIFICE CLASSÉ” is one term.

<sup>17</sup> From Schott (2002).

b. create a coined term in the target language.

*Example:*

| <i>German</i>   | <i>English</i>   |
|-----------------|--|
| BERUFSPRAKTIKUM | VOCATIONAL PRACTICAL<br>SN Use this descriptor for a combination vocational education and practical training |

c. use a broader term in the target language with a qualifier<sup>18</sup>

| <i>German</i> | <i>English</i>                             |
|---------------|--|
| GYMNASIUM     | SECONDARY SCHOOL (GYMNASIUM) <sup>19</sup> |

d. establish one or more non-descriptor(s) in the source language, with a link to the preferred term(s).

*Example:*<sup>19</sup>

| <i>German</i>   | <i>English</i>             |
|-----------------|----------------------------|
| FEUDALISMUS     | FEUDALISM                  |
| MONARCHIE       | MONARCHY                   |
| Feudalmonarchie | USE FEUDALISM and MONARCHY |

In this instance, the last solution is the least artificial.

5. *Non-equivalence:* no existing term with an equivalent meaning is available in the target language for a descriptor in the source language.

This situation is not acceptable in a *symmetric thesaurus*. The simplest solution is the removal of the “orphan” descriptor in the source language, especially if it is a very specialized term.

If removal is not advisable, three options are available.

- Transform the “orphan” term into a non-descriptor and link the non-descriptor to a descriptor with which it shares many essential characteristics.

*Example:*<sup>20</sup>

| <i>English</i>   | <i>French</i>  |
|--|--|
| Classroom environment<br>USE LEARNING ENVIRONMENT                              |  |
| LEARNING ENVIRONMENT<br>UF Classroom environment<br>UF Educational environment | MILIEU EDUCATIF<br>EP Conditions d'apprentissage<br>EP Milieu pédagogique<br>EP Milieu d'apprentissage |

<sup>18</sup> From Schott (2002)

<sup>19</sup> Note that here GYMNASIUM is the German term, in English a ‘gymnasium’ is a room where one does physical exercises.

<sup>20</sup> From Canadian (1992)

- Import the source term into the target language.

*Example*<sup>21</sup>:

| <i>English</i>                                    | <i>French</i>                                   |
|---|---|
| MARKETING<br>RT ADVERTISING<br>RT MARKET RESEARCH | MARKETING<br>TA ÉTUDE DE MARCHÉ<br>TA PUBLICITE |

*Example*<sup>22</sup>

| <i>German</i>     | <i>English</i>  |
|-------------------|---|
| DIAKONISCHES WERK | DIAKONISCHES WERK<br>SN Form of social work done by<br>protestant churches in Germany |

- create an equivalent (a neologism).

*Example*<sup>23</sup>

| <i>German</i> | <i>English</i>           |
|---------------|--------------------------|
| HOCHRECHNUNG  | PROJECTION (STATISTICAL) |

In a *non-symmetric* thesaurus, accept the descriptor in the source without any equivalent in the target language. In the example below ANALPHABÉTISME FONCTIONNEL has no equivalence in English.

*Example*<sup>24</sup>

| <i>French</i>  | <i>English</i>   |
|--|--|
| <b>ALPHABETISME</b><br><b>E</b> : LITERACY<br>EP Lettrisme<br>TS ALPHABETISME FONCTIONNEL<br>TA ALPHABETISATION<br>TA ANALPHABETISME           | <b>LITERACY</b><br><b>F</b> : ALPHABETISATION<br><b>F</b> : ALPHABETISME<br>UF ADULT LITERACY<br>NT ADOLESCENT LITERACY  |
| <b>ALPHABETISME FONCTIONNEL</b><br><b>E</b> : FUNCTIONAL LITERACY<br>EP Semi-alphabétisme<br>TG ALPHABETISME<br>TA ALPHABETISATION FONCTIONNEL | <b>FUNCTIONAL LITERACY</b><br><b>F</b> : ALPHABETISME FONCTIONNEL<br><b>F</b> : ANALPHABETISME FONCTIONNEL<br>UF Functional illiteracy<br>UF Utilitarian literacy<br>BT LITERACY |
| <b>ANALPHABETISME FONCTIONNEL</b><br>TG ANALPHABETISME<br>TA ALPHABETISME FONCTIONNEL  |  |
| <b>ALPHABETISATION FONCTIONNELLE</b><br><b>E</b> : FUNCTIONAL LITERACY   | <b>FUNCTIONAL LITERACY</b><br><i>See above</i>   |

<sup>21</sup> From Canadian (1992). Note that this thesaurus does not contain the term *mercatique*, not even as a non-descriptor.

<sup>22</sup> From Schott (2002)

<sup>23</sup> From Schott (2002)

<sup>24</sup> From Canadian (1992)

## 4 Starting from existing thesauri

### 4.1 Merging

It is possible to build a new multilingual thesaurus by *merging* several thesauri and/or subject heading lists in different languages into one. The purpose is to use the existing vocabularies as much as possible. In practice, this leads to building a new thesaurus based on the vocabularies of the existing ones. In the end the users, both indexers and searchers, have to use the new thesaurus resulting from the merging.

### 4.2 Linking

#### 4.2.1 Introduction

The idea behind *linking* of subject heading languages is that the users, both indexers and searchers, can continue to use the same subject heading language as before. However, through the linking it becomes possible to search in collection A, indexed with subject heading language X, using subject heading language Y, used to index collection B.

Linking can be done with monolingual subject heading languages in the same language, e.g. all built in Russian, with multilingual subject heading languages, and with monolingual subject heading languages in different languages, or with any combination of subject heading languages.

Linking is done by *mapping* headings of subject heading languages. The result can be displayed as a table with as many columns as subject heading languages involved. In each row of the table the “equivalent” terms of the different subject heading languages are given. A cell of the table can contain zero, one or more terms.

#### 4.2.2 Types of equivalence

For mapping, the following level of equivalence can be distinguished:

1. Complete equivalence: in all subject heading languages one, and only one, preferred term exists leading to an equivalent search result.
2. Incomplete equivalence: for each preferred term in any of the subject heading languages, an equivalent term or a syntactical expression<sup>25</sup> in each of the subject heading languages can be found. In at least one subject heading language syntactical expression has to be used.
3. Non-equivalence: for a preferred term in one or more subject heading languages, no equivalent exists in at least one of the other subject heading languages.

##### 4.2.2.1 Complete equivalence

In the case of equivalence, linking may be approximate. Equivalence does not mean here that the terms have the same meaning linguistically, but that it is expected that the

---

<sup>25</sup> A syntactical expression is an expression built up from two or more terms from a subject heading list, e.g. United States–History–Nineteenth century

terms of the different subject heading languages lead to literature about the same subject.<sup>26</sup>

*Example*<sup>27</sup>

| <i>English (LCSH)</i> | <i>French (RAMEAU)</i> | <i>German (SWD)</i> |
|-----------------------|------------------------|---------------------|
| Decathletes           | Décathloniens          | Zehnkämpfer         |
| Decathlon             | Décathlon              | Zehnkampf           |
| Discus throwing       | Lancer du disque       | Diskuswurf          |
| Divers                | Plongeurs              | Kunstspringer       |
| Diving                | Plongeon               | Wasserspringen      |
| Hammer throwing       | Lancer du marteau      | Hammerwurf          |
| Hurdle racing         | Course de haies        | Hürdenlauf          |

#### 4.2.2.2 Incomplete equivalence

In the case of incomplete equivalence, in at least one of the subject heading languages a combination of two or more terms has to be used as equivalents. This combination can be a Boolean AND or a Boolean OR combination.

An OR combination is necessary if for a preferred term in a subject headings language two terms, together with the extension of the first term, have to be used in another subject headings language.

*Example*<sup>28</sup>

| <i>English (LCSH)</i>      | <i>German (SWD)</i>  | <i>French (RAMEAU)</i>                |
|----------------------------|----------------------|---------------------------------------|
| Jumping                    | Hochsprung OR Sprung | Saut en hauteur OR Sauts (athlétisme) |
| Broad jump<br>UF Long jump | Weitsprung           | Saut en longueur                      |

There exists in each of the three subject heading languages a term for broad jumping, but the Library of Congress subject heading JUMPING is used both for ‘jumping in general’ and for ‘high jumping’. So a Boolean OR is needed in the SWD and in RAMEAU when the search starts with subheading JUMPING in the LCSH. A search starting with HOCHSPRUNG in German searches with JUMPING in the catalogue indexed with the LCSH, as well as finding documents about ‘jumping’ in general. But in a catalogue indexed with RAMEAU the exact equivalent SAUT EN HAUTEUR will be used.

An AND combination is necessary when in one subject heading language a term exists that is hierarchically on a lower level than the relevant terms in at least one of the other subject heading languages.

*Example*<sup>29</sup>

| <i>English (LCSH)</i> | <i>German (SWD)</i>    | <i>French (RAMEAU)</i>    |
|-----------------------|------------------------|---------------------------|
| Cycling               | Radspport              | Cyclisme                  |
| Cycling -- Training   | Radspport AND Training | Cyclisme AND Entraînement |

<sup>26</sup> In the MACS project two expressions are ‘approximately equivalent’ when the expressions contain queries that return the most equivalent results possible, given the information stored in both the databases and the SHLs.”

<sup>27</sup> Example from the MACS Project [<http://laborix.uvt.nl/prj/macs/>]

<sup>28</sup> From MACS [<http://laborix.uvt.nl/prj/macs/>]

<sup>29</sup> From Kunz (2002)

To gain access to the literature indexed with *Cycling–Training* using RAMEAU it is recommended to start with *Cyclisme AND Entraînement*. An alternative is that *Cyclisme* gives access to *Cycling* and all combinations starting with *Cycling* -- ....

#### **4.2.2.3 Non-equivalence**

In the case of non-equivalence, there is no access when starting with another subject heading language or thesaurus.

#### **4.2.2.4 Some remarks**

1. Mapping can be done on two levels: between the building blocks of subject heading languages (or thesauri) or between the combinations of building blocks used for subject indexing.

The LCSH heading *Cycling–Training* consists of two building blocks: *Cycling* and *Training*. If this subject heading were mapped to a German subject heading language it could be mapped to *Radsporttraining*, or to the two terms that are equivalent to the building blocks: *Radsport* and *Training*.

2. Mapping on the file level only and not on the application level is recommended. This means that syntactical structures where subject headings are built from building blocks given in an authority file are not mapped.

## **5 Glossary**

### **A-symmetrical thesaurus**

See: Non-symmetrical thesaurus

### **Coined term**

A neologism specially created in a target language to express a concept that is denoted by an existing and recognized term in a source language, but that has not previously been expressed in the target language.

### **Compound concept**

A concept that can be thought of as a combination of two or more other concepts. Examples: comic strip [a *tale* told by means of *drawings* and words in *bubbles*]; trade agreement [an *agreement* about *trade*]. In many languages compound words are used to express a compound concept [English: comic strip, German: Bildergeschichte, Dutch: stripverhaal, French: bande dessinée] [English: trade agreement, German: Wirtschaftsabkommen, Dutch: handelsovereenkomst; French: accord commercial]



### **Compound term**

A compound term is a term consisting of more than one word (a phrase term), or a compound word. Compound words are words composed of more than one word (often nouns), with or without connecting letters.<sup>30</sup>

### **Compound word**

A word that is a combination of two or more others words into one. Examples: Kleuterschool (Dutch, built from the two nouns *kleuter* and *school*); Wirtschaftsabkommen (German, also built from two nouns *Wirtschaft* and *Abkommen*).

### **Concept**

A knowledge unit.

### **Controlled vocabulary**

A limited set of terms or notations in a thesaurus or classification that must be used both for indexing and searching. In a controlled vocabulary consisting of terms taken from a natural language the use of synonyms and homonyms is avoided for terms recommended for use in indexing. In most cases some structure is imposed on the terms and notations so that those whose meanings are related are linked in some way.

### **Descriptor**

A term used consistently in indexing to represent a given concept. Synonym: preferred term

### **Exact equivalence**

The terms representing a particular concept in different languages are identical in meaning and scope and are suitable as preferred terms.

### **Inexact equivalence**

Terms in the languages involved express the same general concept but the meanings are not precisely identical. Often the differences are more cultural than semantic; there is a difference in connotation.

### **Inter-language equivalence**

The terms of a given language have terms in other languages with the same, or partly the same meaning. See also *Exact equivalence*, *Inexact equivalence*, *Non-equivalence*, *Partial equivalence* and *Single-to-many equivalence*.

### **Intra-language equivalence**

Terms within a given natural language are equivalent if they have the same meaning (synonymy) or are treated as terms with the same meaning (pseudo-synonymy).

### **Loan term**

A term in one language (the source language) that has been adopted without change in a second language (the target language)

---

<sup>30</sup> In English different forms of compound terms can exist next to one another: separate words, words connected with hyphens or words written as a compound term.

**Non-descriptor**

A term in a thesaurus that is not used in indexing to represent a given concept, but acts only as an access point that refers to the descriptor that should be used instead. Synonym: non-preferred term

**Non-equivalence**

For a particular term in one language there is no existing term with an equivalent meaning available in other language(s).

**Non-preferred term**

See: Non-descriptor

**Non-symmetrical thesaurus**

A multilingual thesaurus in which the number of descriptors in each language is not necessarily the same and also the way descriptors are related to each other can be different for the different languages.

**Partial equivalence**

The descriptor in one of the languages has a slightly broader or narrower meaning than the descriptor in the other languages.

**Phrase term**

A term that consists of more than one word to express a compound or single concept. Examples: *trade agreement* (English), *accord commercial* (French), *concrete bridges* (English).

**Preferred term**

See: Descriptor.

**Single-to-many equivalence**

To express the meaning of a term in one of the languages more than one term is needed in the other language(s).

**Source language**

The language that serves as a starting point when a term is translated into its (nearest) equivalent term or term(s) in a second (or target) language.

**Subject**

The intellectual content of a document as determined in subject analysis and destined to be translated in descriptors.

**Subject indexing language**

See: Controlled vocabulary.

**Symmetrical thesaurus**

A multilingual thesaurus in which each descriptor has one and only one equivalent descriptor in every language and is related in the same way to other descriptors in the given language.

**Syntactical expression**

An expression built up from two more terms from a subject heading list, e.g. United States–History–Nineteenth century

**Target language**

The language into which a term first encountered in another language (the source language) is translated.

**Thesaurus**

Alphabetically and/or systematically ordered thematic limited inventory of terms between which semantic paradigmatic relationships, such as hierarchy, are established. Synonymy, homonymy and polysemy are controlled as far as possible.

## 6 Literature

**Aitchison (2000)**

Thesaurus construction and use : a practical manual / Jean Aitchison, Alan Gilchrist, David Bowden. – 4<sup>th</sup> ed. – London : Aslib IMI, 2000. – xiv, 218 p. : ill. ; 30cm. – ISBN 0-85142-446-5

**Canadian (1992)**

Canadian literacy thesaurus = Thésaurus canadien d’alphabétisation. – Toronto : Canadian literacy thesaurus Coalition, 1992. – 241, 251 p.

**Dahlberg (1989)**

Concept and definition theory. – In: Classification theory in the computer age: conversations across disciplines. Proceedings from the Conference, November 18-19, 1988, Albany, New York (p. 12-24). Albany, New York: Rockefeller College Press, University at Albany, State University of New York, 1989.

**Hudon (1997)**

Multilingual thesaurus construction : integrating the views of different cultures in one gateway to knowledge and concepts / Michèle Hudon. – In: Information Services & Use 17(1997) 2/3, p. 111-123. – Also in: Knowledge Organization 24(1997) 2, p.84-91

**Hudon (2001)**

Relationships in multilingual thesauri / Michèle Hudon. – In: Relationships in the organization of knowledge / edited by Carol A. Bean and Rebecca Green. – Dordrecht [etc.] : Kluwer Academic Publishers, 2001. – p. 67-80

**ISO 5964:1985**

Documentation - Guidelines for the establishment and development of multilingual thesauri. See: <http://www.iso.org/iso/en/ISOOnline.frontpage>

**Kunz (2002)**

Sachliche Suche in verteilten Ressourcen: ein kurzer Überblick über neuere Entwicklungen. Paper presented by Martin Kunz at the 68th IFLA Council and General

Conference, August 18-24, 2002, Glasgow (UK) 2002. – Also available in English, French and Russian

**Landry (2004)**

Multilingual subject access : the linking approach of MACS / Patrice Landry. – In: *Cataloging and Classification Quarterly* 37(2004)3/4, p. 177-191

**Milstead (2001)**

Standard for relationships between subject indexing terms / Jessica L. Milstead. – In: *Relationships in the organization of knowledge* / edited by Carol A. Bean and Rebecca Green. – Dordrecht [etc.] : Kluwer Academic Publishers, 2001. – p. 53-66

**Nase & Mdivani (1996)**

Creating a multilingual thesaurus for the social sciences : linguistic and intercultural problems / Annemarie Nase, Robert Mdivani. – In: *Social sciences in transition : Social science information needs and provision in a changing Europe : proceedings of as European conference in Berlin, November 11<sup>th</sup>-13<sup>th</sup>, 1994* / edited by Heinrich Best, Ulrike Becker, Arnaud Marks. – Bonn : Informationszentrum Sozialwissenschaften, 1996. – p. 349-356.

**Principles(1999)**

Principles underlying subject heading languages (SHLs) / edited by Maria Inês Lopes and Julianne Beall ; Working Group on Principles Underlying Subject Heading Languages. – München, K. G. Saur, 1999. – x, 183 p. – (UBCIM Publications ; N.S. Vol. 21). – ISBN 3-598-11397-8

**Schott (2002)**

Thesaurus Sozialwissenschaften 2002 / bearbeitet van Hannelore Schott. – Bonn : Informationszentrum Sozialwissenschaften, 2002. – 4 vol.

- Thesaurus Sozialwissenschaften : Alphabetischer Teil. – xx, 332 p. – ISBN 3-8206-0134-6

- Thesaurus Sozialwissenschaften : Systematischer Teil. – xx, 381 p. – ISBN 3-8206-0135-6

- Thesaurus Sozialwissenschaften : Deutsch–Englisch = Thesaurus for the Social Science : German–English. – xvi, 308 p. – ISBN 3-8206-0136-8

- Thesaurus Sozialwissenschaften : Englisch–Deutsch = Thesaurus for the Social Science : English–German. – xvi, 279 p. – ISBN 3-8206-0137-6

**Zeng & Chan (2004)**

Trends and issues in establishing interoperability among knowledge organization systems / Maria Lei Zeng and Lois Mai Chan. – In: *Journal of the American Society for Information Science and Technology* 55(2004)5, p. 377-395

## Appendix A

### *Example of a non-symmetrical thesaurus*<sup>31</sup>

| EDUCATION<br>Éducation |  | ÉDUCATION<br>Education |  |
|------------------------|--|------------------------|--|
| SN                     | Use only for a general discussion of the concept; if possible use a more specific term (Continuing education, etc.)  | SN                     | Utiliser pour une discussion générale du concept d'éducation ; utiliser de préférence un terme spécifique (Éducation communautaire, etc.)  |
| NT                     | Adult education<br>Alcohol education<br>Career education<br>Civic education<br>Community education<br>Compensatory education<br>Consumer education<br>Continuing education<br>Cooperative education<br>Critical education<br>Cultural education<br>Distance education<br>Drug education<br>Elementary education<br>Environmental education<br>Equal education<br>Extension education<br>Formal education<br>General education<br>Health education<br>Informal education<br>International education<br>Law related education<br>Liberatory education<br>Out of school education<br>Political education<br>Popular education<br>Postsecondary education<br>Preschool education<br>Safety education<br>Secondary education<br>Sex education<br>Special education<br>Women's education | NT                     | Éducation antialcoolique<br>Éducation antidroque<br>Éducation communautaire<br>Éducation compensatoire<br>Éducation culturelle<br>Éducation des adults<br>Éducation des femmes<br>Éducation extrascolaire<br>Éducation formelle<br>Éducation internationale<br>Éducation juridique<br>Éducation permanente<br>Éducation politique<br>Éducation populaire<br>Éducation préscolaire<br>Éducation relative à l'environnement<br>Éducation sanitaire<br>Éducation sexuelle |
| RT                     | Access to education<br>Employment education relationships<br>Industry education relationships<br>Instruction<br>Learning<br>Pedagogy<br>Schooling<br>Training<br>Undereducation  | TA                     | Apprentissage<br>Enseignement<br>Formation<br>Pédagogie<br>Rééducation   |

<sup>31</sup> Taken from Canadian (1992).

|   |   |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
|---|---|--|----|-----------|-----------------------------|--|--|----|---|----|--|----|-----------|----|--|
| <p>CONTINUING EDUCATION<br/>Éducation permanente</p> <table border="1" data-bbox="293 667 792 808"> <tr> <td>UF</td> <td>Adult continuing education<br/>Continuing professional education<br/>Ongoing education<br/>University extension</td> </tr> </table> <table border="1" data-bbox="293 835 792 865"> <tr> <td>BT</td> <td>Education</td> </tr> </table> <table border="1" data-bbox="293 892 792 976"> <tr> <td>RT</td> <td>Adult education<br/>Extension education<br/>Professional development</td> </tr> </table> | UF  | Adult continuing education<br>Continuing professional education<br>Ongoing education<br>University extension | BT | Education | RT                          | Adult education<br>Extension education<br>Professional development | <p>ÉDUCATION PERMANENTE<br/>Continuing education</p> <table border="1" data-bbox="824 283 1323 634"> <tr> <td>DF</td> <td>Ensemble d'activités qui assurent, à toutes les époques de la vie, la formation et le développement de la personne, en lui permettant d'acquérir des connaissances, des habiletés ou des comportements, et de développer un ensemble d'aptitudes intellectuelles, manuelles, etc.</td> </tr> </table> <table border="1" data-bbox="824 661 1323 823"> <tr> <td>EP</td> <td>Éducation continue<br/>Éducation postsecondaire<br/>Éducation récurrente<br/>Formation continue</td> </tr> </table> <table border="1" data-bbox="824 850 1323 879"> <tr> <td>TG</td> <td>Éducation</td> </tr> </table> <table border="1" data-bbox="824 907 1323 961"> <tr> <td>TA</td> <td>Cours à l'extension<br/>Éducatifs des adultes</td> </tr> </table> | DF | Ensemble d'activités qui assurent, à toutes les époques de la vie, la formation et le développement de la personne, en lui permettant d'acquérir des connaissances, des habiletés ou des comportements, et de développer un ensemble d'aptitudes intellectuelles, manuelles, etc. | EP | Éducation continue<br>Éducation postsecondaire<br>Éducation récurrente<br>Formation continue | TG | Éducation | TA | Cours à l'extension<br>Éducatifs des adultes |
| UF  | Adult continuing education<br>Continuing professional education<br>Ongoing education<br>University extension  |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| BT  | Education   |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| RT  | Adult education<br>Extension education<br>Professional development  |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| DF  | Ensemble d'activités qui assurent, à toutes les époques de la vie, la formation et le développement de la personne, en lui permettant d'acquérir des connaissances, des habiletés ou des comportements, et de développer un ensemble d'aptitudes intellectuelles, manuelles, etc. |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| EP  | Éducation continue<br>Éducation postsecondaire<br>Éducation récurrente<br>Formation continue  |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| TG  | Éducation   |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| TA  | Cours à l'extension<br>Éducatifs des adultes  |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| <p>CRITICAL EDUCATION</p> <table border="1" data-bbox="293 1129 792 1213"> <tr> <td>DF</td> <td>Provision of information and knowledge necessary to facilitate social change</td> </tr> </table> <table border="1" data-bbox="293 1241 792 1270"> <tr> <td>BT</td> <td>Education</td> </tr> </table>  | DF  | Provision of information and knowledge necessary to facilitate social change                                 | BT | Education | <p>No French equivalent</p> |  |  |    |   |    |  |    |           |    |  |
| DF  | Provision of information and knowledge necessary to facilitate social change  |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |
| BT  | Education   |  |    |           |                             |  |  |    |   |    |  |    |           |    |  |