115

Guidelines for Multilingual Thesauri

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

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

Chairs: Gerhard J. A. Riesthuis (1999-2005) (Netherlands), Patrice Landry (2006-2008) (Switzerland)
Members: Lois Mai Chan (USA), Jonathan Furner (USA), Martin Kunz (Germany), Pia Leth (Sweden), Dorothy McGarry (USA), Ia McIlwaine (United Kingdom),
Max Naudi (France), Marcia Lei Zeng (USA)

Approved by the Classification and Indexing Section
December 12, 2008

Table of Contents

Foreword

Acknowledgements

The Classification and Indexing Section focuses on methods of providing subject access in catalogues, bibliographies, and indexes to documents of all kinds, including electronic documents. The Section serves as a forum for producers and users of classification and subject indexing tools, and it works to facilitate international exchange of information about methods of providing subject access. It promotes standardization and uniform application of classification and indexing tools by institutions generating or utilizing bibliographic records. Prior to developing the *Guidelines for Multilingual Thesauri,* the Section developed *Principles Underlying Subject Heading Languages (SHLs)*, and published the document in 1999.

About the *Guidelines*

The Working Group (WG) on Guidelines for Multilingual Thesauri of the IFLA Classification and Indexing Section was established during the 65th IFLA Congress in Bangkok, Thailand, in August 1999. The WG initiated a project to draft new *Guidelines for Multilingual Thesauri*, to replace the 1976 UNESCO *Guidelines for the Establishment and Development of Multilingual Thesauri*, which were more then 20 years old. The WG has been chaired by Gerhard Riesthuis (University of Amsterdam, The Netherlands) and Patrice Landry (Swiss National Library). Members of the WG are: Lois Mai Chan (USA), Jonathan Furner (USA), Martin Kunz (Germany), Pia Leth (Sweden), Dorothy McGarry (USA), Ia McIlwaine (United Kingdom), Max Naudi (France), and Marcia Lei Zeng (USA). The first draft of the present *Guidelines* was produced in 2002 and a version was submitted for world-wide review in 2005.

Following the world-wide review, a small committee was set up to finalise and edit the *Guidelines* for publication. This group consisted of Lois Mai Chan, Patrice Landry, Dorothy McGarry and Marcia Lei Zeng. The Working Group wishes to thank Jonathan Furner for proofreading the final version of the *Guidelines*.

The objective of this document is to add to the existing guidelines for multilingual thesauri as embodied in the ISO-standard *Guidelines for the Establishment and Development of Multilingual Thesauri* (ISO 5964-1985) and in handbooks on thesaurus building, such as *Thesaurus Construction and Use: A Practical Manual* by Aitchison et al. (2000). The general principles for the building of monolingual thesauri are assumed.

The current *Guidelines* complements other standards for controlled vocabularies such as IFLA's *Principles Underlying Subject Heading Languages (SHLs)* and the American standard ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.*

1 Introduction

Multilingual indexing vocabularies exist in different forms, e.g. subject heading lists, thesauri, enumerative classifications, analytico-synthetic classifications. In a multilingual indexing vocabulary both the terms and the relationships are represented in more than one language.

In this document the emphasis is on multilingual thesauri.

Since the drawing up of the *Guidelines for the Establishment and Development of Multilingual Thesauri* in the 1970s two developments have played important roles in the thinking about multilingual access to information: the building of non-symmetrical thesauri and the linking of two or more thesauri and/or controlled vocabularies.

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) thesaurus to be used in indexing and retrieval
   b. linking existing thesauri and subject heading lists to each other; using the existing thesauri and/or subject heading lists 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[1]. This approach is not discussed 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 indexing vocabularies using any one of the linked thesauri or subject heading lists. An example of a multilingual linking project is the MACS (Multilingual Access to Subjects) project (see http://macs.cenl.org/).

Building from the bottom up is only viable in cases where a new thesaurus or subject heading list is envisaged. The main advantage is that the languages involved can be treated equally.

---

[1] See: Nase & Mdivani (1996) on translation of thesauri.

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

a) Equivalence problems
Semantic problems pertain to equivalence relations between preferred and non-preferred terms in thesauri or subject heading lists. Equivalence relations exist not only within each separate language involved (intra-language equivalence), but also between the languages (inter-language equivalence). Intra-language homonymy and inter-language homonymy are also considered semantic issues. Additional problems pertaining to semantics involve the scope, form and choice of thesaurus terms.

b) 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 indexing vocabularies involved. In other approaches mentioned, it is possible in principle to apply the same structure to all languages. This question will be discussed later.(see § 3.2)

A glossary appears at the end of this document.

2   List of abbreviations (relationship indicators)

The following is a list of relationship indicators used in thesauri to identify a semantic relationship between terms.

Dutch[2]	English	German[3]	French[4]	Meaning
USE	USE	BS	EM	Use term … instead
UF	UF	BF	EP	Use 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

An alternative can be to use the English abbreviations in all language versions of a multilingual thesaurus, as shown here for Dutch

---

[2] In Duttch the English abbreviations are used

[3] The meaning of the German abbreviations is BS: Benutze; BF: Benutzt für; OB: Oberbegriff; UB: Unterbegriff; VB: Verwandter Begriff; D: Definition.

[4] The meaning of the French abbreviations is : EM: Employer; EP: Employé 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..

3 Building a multilingual thesaurus from the bottom up

3.1 Introduction

The morphological aspects, e.g. spelling, of preferred terms and non-preferred terms have been discussed at great length in guidelines for monolingual thesauri[5], in *Principles Underlying Subject Heading Languages (SHLs)*[6] and in the context of the MACS project[7]. 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 subset 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 preferred term must have one and only one equivalent term in every language and be related in the same way to other preferred terms in the given language (a *symmetrical thesaurus*). This can be complete or incomplete equivalence (see 4.2.2). The number of non-preferred terms can be different.

The alternative is a non-identical and non-symmetrical structure where the number of preferred terms in each language is not necessarily the same and also where the way preferred terms 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 preferred terms and relations. This means that each preferred term in any of the languages has an equivalent term in all other languages and that the relations between the preferred terms in all languages are the same. If in language X a generic relation exists between preferred term 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 often does happen, that cross-language equivalences are forced where they do not exist and questionable relational structures occur.

---

[5] For an overview of such guidelines see Milstead (2001).
[6] Principles (1999).
[7] Landry (2004).

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

English (source)	French (target)
Education	Éducation
	Enseignement
	Formation
Teaching	Enseignement
Instruction	Enseignement
	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 preferred terms (see Appendix).

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.

3.3 Morphology and semantics

3.3.1 Scope of preferred terms

Each preferred term included in a thesaurus should represent a single concept[8]. It can be a single-word term or a multi-word compound term. In principle, the preferred terms 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 preferred terms 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 "class-of-one" concepts are expressed as proper nouns. Often official or unofficial versions exist in different languages. A choice has to be made among:

- 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 using the proper name in the original language for the languages that lack it, or
- using translations as far as is sensible and possible.

---

[8] **For clarification on "concept" see Dahlberg (1989).**

In all cases appropriate non-preferred terms should be added. It is advisable to use an appropriate list of authorised names for each language.

English	French	Dutch
International Committee of the Blue Shield (ICBS)	Comité international du Bouclier bleu (CIBB)	Internationaal Comitee van het Blauwe Schild UF International Committee of the Blue Shield (ICBS) UF Comité international du Bouclier bleu (CIBB)

The English and French preferred terms are in official forms of the name of this "class-of-one"; in Dutch the preferred term is a translation with the official English and French terms as non-preferred terms[9].

3.3.2 Clarification and disambiguation of preferred terms

The scope of each preferred term 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 preferred term is placed. Also, the preferred term 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 § 3.3.3) is often sufficient for disambiguation purposes.

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".

---

[9] See http://www.ifla.org/blueshield.htm

English (British)	English (USA)	Dutch	French
Cranes (birds)	Cranes (birds)	Kraanvogels	Grue (oiseau)
Cranes (lifting equipment)	Cranes (lifting equipment)	Hijskranen 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 NT Water taps NT Gas taps	Faucets NT Water faucets NT Gas faucets	Kranen SN voor kranen als hijswerktuig gebruik hijskranen NT Waterkranen NT Gaskranen	Robinet NT Robinet à eau NT Robinet à gaz

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

Dutch	English	German	French
Geschenken UF Giften UF Cadeaus UF Kado's	Gifts UF Presents	Geschenke NT Gaben NT Spenden	Cadeau UF Donation
Vergiften	Poisons	Gifte	Poison

An exception is the case where terms in more than one language are given in one alphabetical order.. This practice is not recommended as a separate list for each language is preferred. Should this option be used, it is recommended to add a language abbreviation in parentheses.

An example from LCSH (English) and RAMEAU (French):
Chance (eng) = Hasard (fre)
Fortune (eng) = Chance (fre)

In the alphabetical index we get:
…
Chance (eng)
Chance (fre)
Fortune (eng)
Hasard (fre)
…

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 preferred term 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, infinitives in some other languages like Dutch or nouns derived from verbs as in Romanian.

Example

English	*Dutch*	*Romanian*
Swimming	Zwemmen	Înot

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

a) adjectival phrases

English	*Dutch*	*Italian*
Concrete bridges	Betonnen bruggen	Ponti in cemento
Library catalogues	Bibliotheekcatalogi	Cataloghi di biblioteca

Note that the phrase "Library catalogues" becomes one (compound) word in Dutch. In Italian the position of the focus and modifier (see 3.3.5) is different from that in English and Dutch.

b) prepositional phrases

English	*Dutch*
Accessories after the fact	Medeplichtigen
Hospitals for children[10]	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 directly equivalent term for "Accessories after the fact". In a symmetrical thesaurus the solution is:

English	*Dutch*
Accessories UF Accessories after the fact	Medeplichtigen

---

[10] This phrase is used here for the sake of the example "Children's hospitals" is more common. Also the Dutch language has "Ziekenhuizen voor kinderen" as a less used equivalent.

In a non-symmetrical thesaurus the solution is:

English	Dutch
Accessories NT Accessories after the fact	Medeplichtigen

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 individual language (recommended in cases where national standards exist) or choose to use the same rules for number in all languages involved.

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

Give all terms in the script of the individual language. Use the appropriate capitalisation, punctuation, diacritics and special characters of the individual language. For spelling, the official rules of the language should be followed. For alphabetisation of terms the rules of the individual language should be used.

The conventions used for the forms of the terms should be clearly explained in the introduction of 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 "Zukunftsangst" that consists of "Zukunft" and "Angst" and means "fear of the future")[11].

Typically, a compound term precoordinates, i.e. pre-combines two or more simple concepts into one compound concept. The use of compound terms in a thesaurus and subject heading list 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 use of compound terms and the use of more precoordination make it more difficult to establish equivalence. A compound term in one language may have an equivalent in another language represented by separate terms.

---

[11] In English different forms of the same compound terms can exist: separate words, words connected with hyphens or words written as a compound term.

The parts of most[12] compound terms can be distinguished as follows:

1. The focus, i.e. the noun component that identifies the general class of concepts to which the term as a whole refers.

Examples:
    a. the noun component "indexes" in the phrase term "Printed indexes",
    b. the noun "hospitals" in the prepositional phrase "Hospitals for children", and
    c. the part "band" in the compound word "Broadband".

2. The 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:
    a. the adjective "printed" in the compound term "Printed indexes",
    b. the preposition-plus-noun combination "for children" in the compound term   "Hospitals for children", and
    c. the part "broad" in the compound word "Broadband".

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 into two or more terms in one language, equivalents in other languages should be split too. If compound terms – especially compound words – are split and are common to one or more languages, a reference by means of a non-preferred term 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 .preferred term. In a non-symmetrical thesaurus, a possibility would be:

English	Dutch
Rail safety	Veiligheid bij spoorwegen USE Veiligheid + Spoorwegen

This possibility has consequences for the structure of the thesaurus (see § 3.2).

---

[12] There are compound terms where the word that is the modifier does not specify a subclass. An example is "Artificial flowers".

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 refers to the hierarchical and associative relations among terms.

Semantic and cultural equivalence is only relevant for preferred terms; in most multilingual thesauri no attempts are made to give equivalent terms for all non-preferred terms. 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 (inter-language synonymy)*[13] *:* the preferred terms in X and Y are semantic and culturally equivalent.

German	*English*	*French*	*Dutch*
Schienennetz	Rail network	Résau ferroviaire	Spoorwegnet
BF Eisenbahnlinie Eisenbahnstrecke	UF Railway line	EP Ligne ferroviaire Voie ferrée	UF Spoorlijn Spoorweg

2. *Inexact or near equivalence (inter-language quasi-synonymy, with a difference in viewpoint):* the preferred terms 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 than semantic, i.e. there is a difference in connotation or appreciation[14]. In the case of inexact equivalence the terms can be treated as if they were exact equivalents.

Example[15]

English	*Spanish*	*French*
Historic settlements	= Asentamientos historicos	≈ Site de peuplement

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

---

[13] The terms in brackets are taken from Eurovoc Version 4.2 (2007). Several examples are taken from existing thesauri. This does not mean approval or disapproval of these thesauri.

[14] Strong-willed and pig-headed both mean stubborn, but the first term has a positive connotation, the second a negative one.

[15] This example is taken from the introduction to the HEREIN Thesaurus (http://www.European-heritage.net/sdx/herein/thesaurus/introduction.xsp) (June 2003).

There are three possible solutions:

a. Treat the terms as exact equivalents.

Example[16]

German	English
Wissenschaft	Science

b. 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[17]

German	English
Wissenschaft	Wissenschaft SN Loan term adopted from German NT Science
Science D Lehnwort aus dem Englishen OB Wissenschaft	Science BT Wissenschaft

c. Treat the situation as *single-to-many equivalence*.(see next case)

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

Two different cases exist:
- First, the target language includes more than one equivalent of the source term (too many target terms).

There are three possible solutions:

a. Create a pre-combined preferred term in the target language.

Example[18]

English	French
Listed building	Édifice inscrit + Édifice classé[19] EP Édifice inscrit EP Édifice classé

---

[16] From Schott (2006).

[17] From ISO 5964-1985, p. 12.

[18] From the HEREIN Thesaurus (see note 15).

[19] Note that "Édifice inscrit" + "Édifice classé" is one term.

b. Modify or specify the source term, e.g. by adding a qualifier.

Example[20]

English	German
Festival (feast) RT Holiday	Fest RT Feiertag
Festival (series of performances) RT Event	Festival RT Event

c. Establish one or more non-preferred term(s) in the target language, with links to the .preferred term

Example

English	German
Festival	Festival BF Fest VB Holiday
	Fest BS Festival

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

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

There are four possible solutions:

a. Create a pre-combined preferred term in the target language.

Example[21]

German	English
Berufspraktikum	Vocational education + Internship

b. Create a coined term in the target language.

Example

German	English
Berufspraktikum	Vocational training SN Use this descriptor for a combination of vocational education and internship

---

[20] From Schott (2006). The qualifiers were added to the terms in this example.
[21] From Schott (2006).

c. Use a broader term in the target language with a qualifier[22].

German	English
Gymnasium	Secondary school (gymnasium) [23]

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

Example[24]

English	German
Town redevelopment	BS Stadt und Sanierung
Town	Stadt
Redevelopment	Sanierung BF Stadtsanierung

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

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

This situation is not acceptable in a symmetrical thesaurus. The simplest solution is the removal of the "orphan" preferred term in the source language, especially if it is a very specialized term.

If removal is not advisable, three options are available.

a. Transform the "orphan" term into a non-preferred term- and link the non-preferred term to a preferred term with which it shares many essential characteristics.

Example:[25]

English	French
Classroom environment USE Learning environment	
Learning environment UF Classroom environment UF Educational environment	Milieu éducatif EP Conditions d'apprentissage EP Milieu pédagogique EP Milieu d'apprentissage

---

[22] From Schott (2006).

[23] Note that here Gymnasium is the German term that refers to a secondary school level in Europe; in English a "gymnasium" is a room where one does physical exercises.

[24] From Schott (2006).

[25] From Canadian (1996).

b. Import the source term into the target language.
    Example[26] :

English	*French*
Marketing RT Advertising RT Market research	Marketing TA Étude de marché TA Publicité

Example[27]

German	*English*
Diakonisches Werk	Diakonisches Werk SN Form of social work done by protestant churches in Germany

c. Create an equivalent new word or phrase.
    Example[28]

German	*English*
Hochrechnung	Projection (statistical)

In a *non-symmetric* thesaurus, the preferred term in the source can be accepted without any equivalent in the target language. For example, in the table below "Analphabétisme fonctionnel" has no equivalence in English.

Example[29]

French	*English*
Alphabétisme E: Literacy EP Lettrisme TS Alphabétisme fonctionnel TA Alphabétisation TA Analphabétisme	Literacy F : Alphabétisation F : Alphabétisme UF Adult literacy NT Adolescent literacy
Alphabétisme fonctionnel E : Functional literacy EP Semi-alphabétisme TG Alphabétisme TA Alphabétisation fonctionnel	Functional literacy F : Alphabétisme fonctionnel F : Analphabétisme fonctionnel UF Functional illiteracy UF Utilitarian literacy BT Literacy
Analphabétisme fonctionnel TG Analphabétisme TA Alphabétisme fonctionnel	
Alphabétisation fonctionnelle E : Functional literacy	Functional literacy *See above*

---

[26] From Canadian (1996). Note that this thesaurus does not contain the term *Mercatique*, not even as a non-descriptor.

[27] From Schott (2006).

[28] From Schott (2006).

[29] From Canadian (1996).

4 Building a multilingual thesaurus 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 resulting new thesaurus.

4.2 Linking/Mapping

4.2.1 Introduction

The idea behind the *linking* of thesauri or subject heading lists is that the users, both indexers and searchers, can continue to use the same subject vocabulary as before. However, through linking it becomes possible to search in collection A, which has been indexed with vocabulary X, using vocabulary Y, which has been used to index collection B. In other words, we can give access to resources indexed through a different thesaurus or subject heading list, using a familiar vocabulary.

Linking can be done: (1) with two or more monolingual subject vocabularies in the same language, e.g. all the linked vocabularies are in Russian; (2) with monolingual subject vocabularies in different languages, e.g. linking a French vocabulary with a German vocabulary; (3) multilingual subject vocabularies; or (4) with any combination of the above.

Linking is usually done by *mapping* terms of thesauri and subject heading lists. The result can be displayed in many ways, such as in a table with as many columns as the number of subject vocabularies involved. In each row of the table the "equivalent" terms of the different vocabularies are given. A cell of the table can contain zero, one or more terms.

4.2.2 Types of equivalence

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

1. Complete equivalence: in all subject vocabularies one, and only one, preferred term exists leading to an equivalent search result.

2. Incomplete equivalence: for each preferred term in any of the vocabularies, an equivalent term or a syntactical expression[30] in each of the vocabularies can be found. In at least one vocabulary a syntactical expression has to be used.

3. Non-equivalence: for a preferred term in one or more subject vocabularies, no equivalent exists in at least one of the other vocabularies.

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 terms of the different subject vocabularies lead to resources about the same subject.[31]

Example[32]

English (LCSH)	French (RAMEAU)	German (SWD)
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 vocabularies a combination of two or more terms has to be used as an equivalent. This combination can be a Boolean AND or a Boolean OR combination (as used in MACS).

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

Example[33]

English (LCSH)	German (SWD)	French (RAMEAU)
Jumping	Hochsprung OR Sprung	Saut en hauteur OR Sauts (athlétisme)
Long jump	Weitsprung	Saut en longeur

---

[30] 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.

[31] 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.

[32] From MACS (http://macs.cenl.org/).

[33] From MACS (http://macs.cenl.org/).

There exists in each of these three subject vocabularies a term for high 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 the heading "Jumping" in LCSH. A search starting with "Hochsprung" in German searches "jumping" in the catalogue indexed with LCSH and finds documents on "high jumping", as well as documents about "jumping" in general. But in a catalogue indexed with RAMEAU the exact equivalent "Saut en hauteur" will be used.

Boolean "AND"
A Boolean AND combination is necessary when in one vocabulary a term exists on a hierarchically lower level than the relevant terms in at least one of the other vocabularies.

Example[34]

English (LCSH)	German (SWD)	French (RAMEAU)
Cycling	Radsport	Cyclisme
Cycling–Training	Radsport AND Training	Cyclisme AND Entraînement

For example, to gain access to the literature indexed with "Cycling–Training" using RAMEAU, it is recommended that the terms "Cyclisme AND Entraînement" be used to start a search query. 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, the user should be informed that there is no equivalent term in the other languages(s).

4.2.2.4 Some remarks
1. Mapping can be done on two levels: between the building blocks of lists of subject headings 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 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.

---

[34] From Kunz (2002).

5 Glossary

Asymmetrical thesaurus
    See: Non-symmetrical thesaurus.

Coined term
    A new word or phrase 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 term consisting of more than one word, or a compound word, that represents a single concept.

Compound word
    A word that is a combination of two or more 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; something that can be understood, thought or meant.

Controlled vocabulary
    A limited set of terms or notations in a thesaurus, subject heading list or classification scheme 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
    See: Preferred term

Exact equivalence
    The relationship between terms representing a particular concept in the same language or different languages that are identical in meaning and scope and are suitable as preferred terms. Within one language one of the terms should be chosen as the preferred term, and the other terms can be included as non-preferred terms.

Facet
A grouping of concepts of the same inherent category. Examples of categories that may be used for grouping concepts into facets are: activities, disciplines, people, materials, places, etc. (ANSI/NISO Z39.19-2005).

Focus
In a compound term, the noun component that identifies the class of concepts to which the term as a whole refers. (ANSI/NISO Z39.19-2005). See also: Modifier.

Homograph
One of two or more words that have the same spelling, but different meanings and origins. In controlled vocabularies, homographs are generally distinguished by qualifiers. (ANSI/NISO Z39.19-2005).

Homonymy
The quality of a word or term having two or more different meanings.

Inexact equivalence
The relationship between terms that express the same general concept but whose meanings are not precisely identical. Often the differences are more cultural than semantic; there is a difference in connotation.

Inter-language equivalence
The relationship between terms in different languages that have 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
The relationship between terms within a given natural language that have the same meaning (synonymy) or are treated as terms with the same meaning (pseudo-synonymy).

Linking
See: Mapping

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

Mapping
A set of correspondences between categories, schema element names, or controlled terms. Mappings are used for transforming data or queries from one vocabulary for use with another. (ANSI/NISO Z39.19-2005)

Modifier
In a compound term, one or more components that serve to narrow the extension of the focus and specify one of its subclasses. Also known as: Difference.

Non-descriptor
    See: Non-preferred term.

Non-equivalence
    Given a particular term in one language, the non-existence in another language of a term with an equivalent meaning.

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

Non-symmetrical thesaurus
    A multilingual thesaurus in which the number of preferred terms in each language is not necessarily the same. The way preferred terms are related to one another can also be different for the different languages.

One-to-many equivalence
    The relationship between a given term in one language and the multiple terms that are needed to express its meaning in another language.

Partial equivalence
    The relationship between two terms, one of which has a slightly broader or narrower meaning than the other.

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).

Polysemy
    See: Homonymy.

Preferred term
    Term used to represent a concept when indexing.
    Note: Preferred terms are sometimes known as descriptors in a thesaurus.

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 into preferred terms..

Subject heading
    A word or phrase, or any combination of words, phrases, and modifiers, used to describe the topic of a content object. Precoordination of terms for multiple and related concepts is a characteristic of subject headings that distinguishes them from controlled vocabulary terms. (ANSI/NISO Z39.19-2005).

Subject heading list

An alphabetical list of subject headings with cross-references from non-preferred terms and links to related terms. These lists often include separate sequences of standardized subheadings that may be combined with all or only some subject headings. (ANSI/NISO Z39.19-2005).

Subject indexing language

See: Controlled vocabulary.

Symmetrical thesaurus

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

Synonymy

The quality of a word or term having exactly or very nearly the same meanings as another word or term.

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.

Term

One or more words designating a concept.

Thesaurus

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

6 References

Aitchison et al. (2000)

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

ANSI/NISO Z39.19-2005

Guidelines for the construction, format, and management of monolingual controlled vocabularies. – Bethesda : NISO, 2005. http://www.niso.org/standards/standard_detail.cfm?std_id=814

Canadian (1996)

Canadian literacy thesaurus = Thésaurus canadien d'alphabétisation. – 2nd ed. – Toronto : Canadian Literacy Thesaurus Coalition, 1996. http://thesaurusalpha.org /thesaurus/about.htm

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.

Eurovoc (2007)

Eurovoc thesaurus -- Version 4.2 / European Communities. – Brussels : EU, 2007. http://europa.eu/eurovoc

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 : Kluwer Academic Publishers, 2001. – p. 67-80.

ISO 5964-1985

Documentation - Guidelines for the establishment and development of multilingual thesauri. – 1st ed. / prepared by the Technical Committee ISO/TC 46, Documentation. – [Geneva] : ISO, 1985. – iii, 61 p. 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 & 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 : 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 a European conference in Berlin, November 11th-13th, 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. New series, vol. 21). – ISBN 3-598-11397-8

Schott (2006)

Thesaurus Sozialwissenschaften = Thesaurus for the social sciences 2006 / bearbeitet van Hannelore Schott. – Bonn : Informationszentrum Sozialwissenschaften, 2006. – 4 vol.
- Thesaurus Sozialwissenschaften : Alphabetischer Teil. – xix, 346 p. – ISBN 3-8206-0151-1
- Thesaurus Sozialwissenschaften : Systematischer Teil. – xix, 396 p. – ISBN 3-8206-0151-1
- Thesaurus Sozialwissenschaften : Deutsch–Englisch = Thesaurus for the social sciences : German–English. – xv, 313 p. – ISBN 3-8206-0154-6
- Thesaurus Sozialwissenschaften : Englisch–Deutsch = Thesaurus for the social sciences : English–German. – xv, 291 p. – ISBN 3-8206-0155-4

UNESCO (1976)

Guidelines for establishment and development of multilingual thesauri. – Paris : Unesco, 1976. – 49 p.

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 & Technology 55(2004) 5, p. 377-395.

Appendix

Example of a non-symmetrical thesaurus[35]

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

---

[35] Taken from Canadian (1996).

			NA	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.
UF	Adult continuing education Continuing professional education Ongoing education University extension		EP	Éducation continue Éducation postscolaire Éducation récurrente Formation continue
BT	Education		TG	Éducation
RT	Adult education Extension education Professional development		TA	Cours à l'extension Éducations des adultes Formation des adultes

CRITICAL EDUCATION		No French equivalent
DF	Provision of information and knowledge necessary to facilitate social change	
BT	Education	