Guidelines for Multilingual Thesauri

Patrice Landry
Chair (2006-2008)
WG on Guidelines for Multilingual Thesauri
Swiss National Library
Bern, Switzerland

Meeting: **215 . Division IV – Bibliographic Control**

Abstract:

In 2008, the IFLA Classification and Indexing Section approved the "Guidelines for Multilingual Thesauri" as submitted by its Working Group and these Guidelines were published in 2009 in the IFLA Professional report series. The work undertaken since 1999 by the section's Working Group on Guidelines for Multilingual Thesauri aimed at developing guidelines that would support and complement other standards work in the area of multilingual thesauri and other subject controlled vocabularies. In particular, the Guidelines examined two developments that occurred since the 1970s in the area of multilingual access to information: the building of non-symmetrical thesauri and the linking of two or more thesauri and /or controlled vocabularies.

The Working Group on Guidelines for Multilingual Thesauri of the IFLA Classification and Indexing Section was established in 1999 in order to look at a growing problem in multilingual subject access. Since the beginning of that decade, with the growth of online catalogues and specialized subject databases available on the internet, the

issue of ensuring multilingual subject access across these resources led to many multilingual thesauri initiatives. At that time, the traditional method of constructing a multilingual thesauri was the translation of a thesaurus into one or more languages as supported by the ISO 5964-1985 standard. In this standard, the languages of a multilingual thesaurus have equal status when each descriptor has one and only one equivalent in a target language and is related to the same other descriptors. The issue of non-equivalence is solved by using "coined terms" where the target language has no equivalent term or when the source language has two equivalences in the target languages. This approach aims at full correspondence between preferred terms and relations. This means that each preferred term in any of the languages and all of the relations between the preferred terms in all languages are the same. It is then assumed that each language has equal status. In this approach, all the different languages of a multilingual thesaurus are considered identical and symmetrical.

The mandate of the WG was to examine the issue of validity of building multilingual thesauri and multilingual subject heading language lists that are not necessarily identical and symmetrical. The WG had been aware of translation and vocabulary linking projects whose aim was building multilingual access tools that attempted to merge two or more existing thesauri into a new one or linking existing thesauri and subject headings list to each other to be used for indexing and/or searching. The usual problems faced were how to integrate a language in another language with a culturally different conceptual structure or trying to fit one conceptual language structure from one culture to another one.

The WG led by Gerhard Riesthuis took up this challenge to draft new guidelines for multilingual thesauri to replace the existing *Guidelines for the establishment and development of multilingual thesauri* or at the very least, to influence standard bodies in incorporating its proposals in future revisions of the multilingual standard. The members of the working group were: 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 Guidelines was produced in 2002 and a version was submitted for world-wide review in 2005. Following this review, a smaller committee was set up to finalize and

edit the Guidelines for publication. This group consisted of Lois Mai Chan, Patrice landry, Dorothy McGarry and Marcia Lei Zeng. Patrice landry chaired the working Group during this final stage of the project (2006-2008). The Guidlelines were published in 2009 in the IFLA Professional Reports Series (also available at: http://archive.ifla.org/VII/s29/pubs/Profrep115.pdf .

The present Guidelines address two approaches in the development of multilingual thesauri / subject heading lists, namely:

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

Combining existing thesauri:

a) Merging two or more existing thesauri into a new multilingual thesaurus

b) Linking existing thesauri and subject heading lists to each other

For both approaches, the Guidelines deal with particular issues and problems encountered in multilingual thesauri construction and linking. A particular focus is given to the equivalence and structural problems where all elements of a thesaurus are explained using examples from existing thesauri.

In the section devoted to the building of a multilingual thesaurus from the bottom up, the guidelines focus mainly on developing a non-symmetrical thesaurus. In regard to the aspects of morphology and semantics, explanations are given on how to deal with the scope and clarification of preferred terms, the use of homographs and qualifiers, forms of terms and compound terms. In many cases, solutions are proposed on how to establish equivalences with particular form of terms in a symmetrical and non-symmetrical thesaurus. The other section refers to the aspect of inter-language equivalence: semantic, cultural and structural. It principally deals with the semantic and cultural aspects as they relate to preferred terms. The structural aspect that refers to the hierarchical and associative relations among terms is not described in the Guidelines. This practice of establishing equivalence for non-preferred terms is not a common practice in multilingual thesauri.

The section on building a multilingual thesaurus starting from existing thesauri is principally concerned about linking / mapping practices. The central focus of linking of thesauri or subject heading lists is the users, both indexers and searchers, who can use a term from their vocabulary that has been linked to a term from another vocabulary to search the collection indexed by the linked term. Linking can be done using different types of vocabularies; two or more vocabularies from the same language, from different languages and from different types of vocabularies (thesauri and subject headings languages). The Guidelines use examples from the MACS (Multilingual Access to Subjects) project to describe the different types of equivalences (complete, incomplete and non equivalence).

A glossary is also included in the Guidelines.

The Guidelines complement other standards for controlled vocabularies such as the IFLA's Principles Underlying Subject heading languages (SHLs), the American ANSI/NISO Z39.19 2005 Guidelines for the Construction, Format and management of Monolingual Controlled Vocabularies as well as the British BS 8723-4:2007 Structured vocabularies for information retrieval – Guide Part 4: Interoperability between vocabularies. Since 2002, the Guidelines were also used by linking projects, such as MACS to validate linking practices and principles.

During the Milan Conference, the Classification and Indexing Section will discuss the present Guidelines in view of setting up a new working group to further develop and improve the current document.