
International Standard



5964

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION • МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ • ORGANISATION INTERNATIONALE DE NORMALISATION

Documentation — Guidelines for the establishment and development of multilingual thesauri

Documentation — Principes directeurs pour l'établissement et le développement de thesaurus multilingues

First edition — 1985-02-15

UDC 025.48

Ref. No. ISO 5964-1985 (E)

Descriptors : documentation, subject indexing, information retrieval, thesauri, multilingual thesauri, preparation, rules (instructions).

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council. They are approved in accordance with ISO procedures requiring at least 75 % approval by the member bodies voting.

International Standard ISO 5964 was prepared by Technical Committee ISO/TC 46, *Documentation*.

Contents

	Page
0 Introduction	1
1 Scope and field of application	1
2 References	2
3 Definitions	2
4 General	3
5 Abbreviations and symbols	4
6 Vocabulary control	5
7 The establishment of a multilingual thesaurus: general problems	5
8 The establishment of a multilingual thesaurus: management decisions	6
9 The establishment of a multilingual thesaurus: language problems	7
10 Establishing equivalent terms in different languages	11
11 Other language problems	20
12 Relationships between terms in a multilingual thesaurus	25
13 Display of terms and relationships	27
14 Form and contents of a multilingual thesaurus	31
15 Organization of work	32
 Annexes	
A Symbolization of thesaural relationships	35
B Examples of displays	36

This document is a preview generated by EVS

This page intentionally left blank

Documentation — Guidelines for the establishment and development of multilingual thesauri

0 Introduction

A trend towards the international exchange of information, fully supported by the UNISIST* programme of UNESCO, and exemplified by systems like the International information system for the agricultural sciences and technology (AGRIS) and the International Nuclear Information System (INIS), clearly calls for a higher commitment to multilingual cooperation. Information systems are expanding across language boundaries, leading to a notable increase in the provision of indexing and retrieval tools which are either language-independent (the Broad System of Ordering), or multilingual. Aids of this kind are essential if retrieval of documents indexed in more than one language is not to depend on the acquisition and use of a single, dominant language. Indexers or searchers should, where possible, be able to work in their mother tongues, or at least in a language with which they are already familiar. Within this context it is considered that multilingual thesauri have a significant part to play in improving the bibliographic control of literature on a global scale.

The standardization of procedures for the construction of a multilingual thesaurus is seen as a primary step in achieving compatibility between thesauri produced by indexing agencies using terms selected from different natural languages. The recording of these procedures will also enable indexers engaged in this task to benefit from the experience of others, and to work in a logical and consistent fashion, using recommended practices which have been established in the course of discussions at an international level.

1 Scope and field of application

1.1 The guidelines given in this International Standard should be used in conjunction with ISO 2788, and regarded as an extension of the scope of the monolingual guidelines. It is considered that the majority of procedures and recommendations contained in ISO 2788 are equally valid for a multilingual thesaurus. This applies particularly to general procedures concerning, for example, the forms of terms, the basic thesaural relationships, and management operations such as evaluation and maintenance. Except when it appears to be necessary, the procedures described in ISO 2788 are not repeated here, and it is therefore essential to refer to both of these International Standards when constructing a multilingual thesaurus.

1.2 These guidelines are restricted in scope to the problems of multilingualism which can arise during the construction of a "conventional" thesaurus, i.e. a thesaurus displaying terms selected from more than one natural language, these terms then constituting the vocabulary of a controlled indexing language. Throughout this International Standard, a distinction is made between *preferred terms* and *non-preferred terms* (see definitions in clause 3). These guidelines are not applicable to indexing languages in which concepts are expressed entirely as symbols (for example mathematical equations or chemical formulae), nor to systems which are based on the automatic analysis and searching of free text. It is considered, however, that a well-constructed multilingual thesaurus can play a significant part in improving retrieval from a free-text system which covers documents in more than one language.

1.3 Multilingual thesauri are relatively recent developments in the field of documentation, and it is inevitable, therefore, that the present guidelines should display certain limitations.

- a) The examples used to illustrate problems encountered in the establishment of term equivalences have been drawn largely from the fields of science (including the social sciences) and technology. As far as possible, however, examples were chosen which illustrate general problems and procedures, i.e. those which should apply in any field of knowledge.
- b) It is realized that the procedures described in these guidelines may not be entirely appropriate for all languages. The examples have been selected, for entirely pragmatic reasons, from three of the major languages, i.e. English, French and German, but this does not imply that these languages are regarded as dominant in the field of documentation. As far as possible the procedures considered here, together with their accompanying examples, relate to problems which may be encountered in any language.

* Intergovernmental programme for co-operation in the field of scientific and technological information.

2 References

ISO/R 639, *Symbols for languages, countries and authorities*.

ISO 1086, *Documentation — Title-leaves of a book*.

ISO 2788, *Documentation — Guidelines for the establishment and development of monolingual thesauri*.

3 Definitions

For the purposes of this International Standard, the following definitions apply:

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

3.2 compound term: An indexing term (see 3.8) which can be factored morphologically into separate components, each of which could be expressed, or re-expressed, as a noun that is capable of serving independently as an indexing term.

NOTE — The parts of the great majority of compound terms can be distinguished as follows:

- a) the **focus** or **head**, i.e. the noun component which identifies the general class of concepts to which the term as a whole refers;
- b) the **difference** or **modifier**, i.e. one or more further components which serve to narrow the extension of the focus by specifying one of its subclasses.

In French, English and similar languages, compound terms usually consist of separate words, whereas the same concept would frequently be expressed by a single word in German and some other languages.

Examples:

a) English		German
SYSTEMS ANALYSIS	=	SYSTEMANALYSE
b) French		German
PONT EN BÉTON	=	BETONBRÜCKE

In example (a) the English word "analysis" and the German component "analyse" both represent foci, and the modifying differences are represented by "systems" (English) and "system" (German). Despite these surface structural differences, however, the terms "systems analysis" and "Systemanalyse" are both regarded as compound terms for the purposes of this International Standard.

3.3 dominant language: An exchange language (see 3.5) which is also used for indexing and retrieval in systems which, for policy reasons, do not give equal status to all the languages in the system. Every concept recognized in the system must necessarily be represented by a preferred term in the dominant language. In some cases, however, an equivalent expression may be lacking in one more of the other languages. These other languages are then known as *secondary languages*.

3.4 equal status: Languages in a multilingual thesaurus have equal status when every preferred term in one language is matched by an equivalent preferred term in all other languages.

3.5 exchange language: The language used as a medium for data exchange in those multilingual systems which, as a matter of policy, decide to use terms selected from only one language for this purpose. The exchange language may also be used for indexing and/or retrieval, and the multilingual thesaurus then functions principally as a means for translating the local languages of indexers and enquirers into, or out of, the exchange language. The different languages in such a system would still be recognized as having equal status (see 3.4) if equivalents are established reciprocally between the preferred terms in the exchange language and the preferred terms in all other languages.

3.6 feedback: The act of changing the form or structure of a term in a source language in order to achieve an easier or a more useful solution to a problem encountered in a target language.

Example: Let us assume that a German thesaurus is used as a source language and contains the term "Lehrerbildungsgesetz". Direct translation of this term into English or French would call for a complicated paraphrase, "Law of education of teachers", or "Loi sur la formation des enseignants". Neither of these phrases would be regarded as a satisfactory indexing term. A shorter expression, which is closer to the German construction, can be achieved in English, i.e. "Teacher education law", but this cannot be done in French.