

---

# UNIT 15 VOCABULARY CONTROL: SUBJECT HEADING LISTS AND THESAURI

---

## Structure

- 15.0 Objectives
- 15.1 Introduction
- 15.2 Vocabulary Control
  - 15.2.1 Meaning and Need
  - 15.2.2 Objectives
  - 15.2.3 Requirements
  - 15.2.4 Entry Vocabulary and Index Vocabulary
- 15.3 Subject Headings list
  - 15.3.1 Meaning
  - 15.3.2 List of Subject Headings - General Principles
- 15.4 Subject Authority File
  - 15.4.1 Meaning
  - 15.4.2 Levels of Subject Authority Control
  - 15.4.3 Functions of a Subject Authority File
- 15.5 Thesaurus
  - 15.5.1 Definition
  - 15.5.2 Functions of a Thesaurus
  - 15.5.3 Types of Thesauri
  - 15.5.4 Structure of Thesaurus
  - 15.5.5 Relationships Between Terms
  - 15.5.6 Thesauri and Subject Headings List
  - 15.5.7 Thesauri and Classification Schemes
  - 15.5.8 Thesaurus-facet
  - 15.5.9 Classaurus
  - 15.5.10 Role of Thesaurus in Indexing and Searching
  - 15.5.11 How to Use a Thesaurus
  - 15.5.12 Thesaurus in Post-coordinate Indexing Systems
- 15.6 Summary
- 15.7 Answers to Self Check Exercises
- 15.8 Key Words
- 15.9 References and Further Reading

---

## 15.0 OBJECTIVES

---

The concept of subject catalogue that provides subject approach to the contents of documents in libraries through the techniques of subject indexing have been discussed in Unit 14 of this Block. In this Unit, you are introduced to the concept of vocabulary control.

After reading this Unit, you will be able to:

- explain the meaning, need and objectives of vocabulary control;
- distinguish controlled vocabulary from natural language vocabulary;
- describe the main features of vocabulary control devices, such as subject headings list and thesauri; and
- discuss their applicability in libraries for subject indexing purposes.

---

## 15.1 INTRODUCTION

---

The use of controlled vocabularies has been an essential element in librarianship for at least a century. A classification scheme, as is used to arrange books on shelves and entries in a classified



is one form of subject headings. We shall now discuss the basic principles of subject headings as adopted by two important standard lists of subject headings (i.e. LCSH and Sear's List) in this Unit to give a proper perspective to the discussion on construction and maintenance of subject headings. Libraries use controlled terminology along with the cross-references. This is accomplished through a control system, called 'subject authority system', which, for each term, records the basis for decisions on the term and what links connect it with other terms. Subject authority systems are discussed here to get an idea of the ongoing maintenance of the subject headings list. The concept of 'thesaurus' is a relatively recent introduction in the area of structure, hierarchical and non-hierarchical relationships between terms and the ways and means by which it differs from other vocabulary control devices are discussed here. Other forms of thesaurus, such as Thesauro-facet and Classaurus have also been discussed in this Unit. The use of a thesaurus in indexing and searching has also been explained.

---

## **15.2 VOCABULARY CONTROL**

---

You have learnt that subject cataloguing or subject indexing are the process used for describing the subject matter of documents. Subject cataloguing/indexing involves assigning terms to represent what the document is about. The complete set of index terms in a subject indexing system may be referred to as the vocabulary or index language of that system. In the case of most libraries, documents are arranged on shelves while the index is likely to comprise entries within a card catalogue representing these documents and providing access to them under the selected Index terms to represent their subject matter. In other situations, the index will be in machine-readable form (on magnetic tape or disk), or microfilm, or in printed book form. In most cases the vocabulary used will be a controlled vocabulary i.e., a limited set of terms that must be used by indexers and searchers. The controlled vocabulary is most often used to standardize descriptors or subject headings representing the contents of documents or subject interest profiles of users (i.e., search strategies and users' profile used in SDI). In principle, however, such a tool can be applied in situations in which the standardization of terminology is needed. In this section, we shall study some of its important features.

### **15.2.1 Meaning and Need**

The term 'Vocabulary control' refers to a limited set of terms that must be used to index documents, and to search for these documents, in a particular system. It may be defined as a list of terms showing their relationships and used to represent the specific subject of a document. A certain degree of structure is introduced in a controlled vocabulary so that terms whose meanings are related are brought together or linked in some way. An uncontrolled vocabulary, is an unlimited set of terms drawn from natural language and used for describing the contents of documents. A natural language is 'natural' in the sense that it grows freely, free from any control whatsoever. Therefore, it is hardly possible to keep a natural language clear of synonyms and homonyms. If we use natural language for subject indexing, subject matter may be described by any one of the Words or phrases, without limitation, such as those occurring in documents themselves. However, certain problems in scotching do arise when no control is imposed on the vocabulary. This is because of the fact that a natural language contains a large number of synonyms, quasi-synonyms, homonyms, acronyms, ambiguous terms, etc. Hence, if vocabulary control is not exercised, different indexers or the same indexer might use different terms for the same concept on different occasions for indexing the documents dealing with the same subject and also use a different set of terms for representing the same subject at the time of searching. This, in turn, it would result in a 'mis-match' and thus affect information retrieval. Thus, the need exists for vocabulary control. In short, we can say that the need for vocabulary control arises overcome -following problems:

- a) occurrence of imprecisely defined words.
- b) rapidly changing terminology.
- c) numerous synonyms for a term. and
- d) problem of homographs.

### **15.2.2 Objectives**

There are basically two objectives for having a controlled vocabulary:

- a) to promote the consistent representation of the subject matter of documents by indexers



and searchers, thereby avoiding the dispersion of related documents, through control of synonymous and nearly synonymous expressions and by distinguishing among homographs; and

- b) to facilitate the conduct of a comprehensive search, by bringing together in someway, the terms that are most closely related semantically.

The first of these objectives is achieved by controlling the terminology in various ways. First, the form of term is controlled, whether this involves grammatical form, spelling, singular and plural form, abbreviations or compound form of terms. Second, a choice is made between two or more synonyms, near-synonyms and quasi-synonyms. Third, homographs are distinguished. The control of synonyms is achieved simply by choosing one of the possible alternatives as the 'preferred term' and referring to this term (by using 'see' or 'use' references) from the variants under which certain users may be likely to approach. It should be obvious that the synonym selected as the preferred term (i.e. the searched for) must be the one under which the majority of users are likely to look first. Sometimes 'quasi-synonyms' are treated in the same way as synonyms (i.e. one is chosen and the reference is made from the other). The term 'quasi-synonyms' is not very precise. Many authors consider the quasi-synonyms as the antonyms that represent opposite extremes on continuum values. An example is the pair of word - 'roughness' and 'smoothness'. Clearly, 'roughness' may be regarded as merely the 'absence of smoothness', and *vice versa*.

The controlled vocabulary also distinguishes among homographs (i.e. words with identical spelling but different meanings), usually by means of a parenthetical qualifier or scope note. Thus CRANE (Bird) tells us that the term is to be used exclusively for a type of birds and not as a lifting equipment or any other possible context.

By controlling synonyms, near synonyms and quasi-synonyms, and by distinguishing among homographs, the vocabulary control device avoids the dispersion of like subject matter and the collocation of unlike subject matter. In this way, it helps to achieve the objectives of consistent representation of subject matter in indexing and searching.

The second objective of vocabulary control is to link together terms that are semantically related in order to facilitate the conduct of comprehensive searches. For a hierarchically related (in a formal genus-species relationship) term, it will also reveal semantic relationship across hierarchies.

### 15.2.3 Requirements

The important requirements of a controlled vocabulary are as follows:

- a) It should have 'warrant' derived from the terminology of literature and the information needs of the actual or potential users. That is to say, a term is justified (warranted) if (i) literature on this subject is known to exist, and (ii) requests for information on this subject (denoted by the term in question) are likely to be made fairly frequently.
- b) It must be sufficiently specific to allow the conduct of the great majority of searches at an acceptable level of precision. This implies that the level of specificity will vary over the vocabulary, some subject areas being developed in greater detail than others. A 'vocabulary developed by National Library of Medicine would need only a few general terms in Mathematics while one developed by the American Mathematical Society would need only a few general terms of a medical nature.
- c) It should be sufficiently pre-coordinate to avoid most problems of false coordination and incorrect term relationships. One way of achieving this, and also of economizing on the absolute size of the vocabulary, is through the use of subheadings.
- d) It should promote consistency in indexing and searching by the control of synonyms, near synonyms, and quasi-synonyms.
- e) It should reduce terminological ambiguity through the separation of homographs and through the definition of terms whose meaning or scope would otherwise be unclear.
- f) It should assist the indexer and searcher in the selection of the most appropriate terms needed to represent a particular subject concept through its hierarchical and cross-reference structure.

### 15.2.4 Entry Vocabulary and Index Vocabulary

Terms linked by equivalence relationships are rather different from those in the other two groups (hierarchical and associative relationships). Here we select one preferred term and use only that

in our index, whereas the other two kinds of relationships occur between terms which are both used in the index. In our indexing language we will, therefore, have both: the preferred terms, which are used for indexing, and the non-preferred terms, which are not. The preferred terms on their own form the 'index vocabulary', while the preferred and the non-preferred terms together form the 'entry vocabulary'. The entry vocabulary is very important. There will be many occasions when we decide, for one reason or another the 'not to use a particular term', but to use the one already in the index vocabulary, instead. The terms in the entry vocabulary should reflect not only literary warrant but also enquiry warrant. In other words, not only those terms found in the literature, but also those used by readers looking for information. We must be aware of the terms used by the users of our information retrieval system as well as those used by the authors whose works we are indexing.

**Self Check Exercises**

- 1) What do you understand by 'vocabulary control'?
- 2) State the objectives of vocabulary control.

**Note:** i) Write your answers in the space given below.  
ii) Check your answers with the answers given at the end of this Unit.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

---

**15.3 SUBJECT HEADINGS LIST**

---

Subject heading has been defined as a word or group of words indicating a subject under which all materials dealing with same theme is entered in a catalogue or bibliography, or is arranged in a file. Credit should go to Crestadoro who, for the first time in his book *The Art of Making Catalogues* published in 1856, could realize that the cataloguer should provide a standardized guide to the subject content of a book by giving it a heading. In 1895, the first standard list of subject headings appeared. The 'List of Subject Headings for Use in Dictionary Catalogues', produced by a committee of the American Library Association (ALA) of which C.A. Cutter was a prominent member, was based on Cutter's principles. It went through three editions (1895, 1898, 1911). In 1910-1914, when Library of Congress (LC) began publishing its list under the title "Subject Headings Used in the Dictionary Catalogs of the Library of Congress", it was found



unnecessary to continue the ALA list. A printed list of subject headings incorporates the thought and experience of many librarians of various types of libraries.

### **15.3.1 .Meaning**

A vocabulary control device depends on a master list of terms that can be assigned to documents. Such a master list of terms is called 'List of Subject Headings'. A list of subject headings list contains the subject access terms (preferred terms) to be used in the cataloguing or indexing operation at hand. When there are synonymous terms for a given subject, these terms are included in the list as these direct the searcher to the preferred terms for the subject. The links from non-preferred terms are called "see" references, and the links to related terms are called "see also" references. This is accomplished through a control system, called 'subject authority system', which, for each term, documents the basis for decisions on the term and on what links connect it with other terms.

### **15.3.2 List of Subject Headings - General Principles**

The rules for subject headings in a dictionary catalogue were formulated by Charles Ammi Cutter in 1876 in his 'Rules for a Dictionary Catalog'. These rules formed the basis of subject headings in American libraries for years to come and are a strong force even today. In respect of subject cataloguing, Cutter stated two objectives:

- a) to enable a person to find a book of which the subject is known, and
- b) to show what the library has on 'a given subject.

The first objective refers to the need to locate individual items, and the second refers to the need to collocate materials on the same subject. It was on the basis of these needs that Cutter set forth his basic principles of subject entry. They are important because the impact of his principles on construction and maintenance of subject headings is still discernible today. Both, the LCSH and the SLSH adopted the Cutter's principles in assigning subject headings for a document. These are discussed in Unit 16 of this Block. The general principles that guide the indexers in the choice and rendering of subject headings from the standard lists of subject headings are discussed in the following sub-sections.

#### **Specific and Direct Entry**

The principles of specific and direct entry require that a document be assigned directly under most specific subject heading that accurately and precisely represents its subject content. If a document is about penguins, it should be entered directly under the most specific heading, that is 'Penguins', it should not be entered under the heading 'Birds' or even under 'Water Birds'. If the name of a specific subject is not available in the list of subject headings, a broader or more general heading may be used. In such cases, the broader heading is the most specific authorized heading in the hierarchy that covers the content of the work. In many cases, several headings may be assigned in order to cover different aspects of a subject.

#### **Common Usage**

This principle states that the word(s) used to express a subject must represent common usage. This means that current American spelling and terminology 'Labor', 'Elevators' not 'Lifts' are to be used. In British libraries these choices would be reversed. According to popular principle, subject headings are to be chosen keeping in mind the needs of the users who are likely to use the catalogue. The popular or common name may be best suited to the readers of a small public library, but in a research or special library, scientific or technical name as heading might be more appropriate. After deciding on the name of heading, the cataloguer should make a reference from the unused heading to the form used. Such references will be discussed latter under Cross References.

#### **Uniformity**

The principle of uniform heading is adopted in order to show what the library collection has on a given subject. One uniform term must be selected from several synonyms, and this term must be applied consistently to all documents on the topic. The heading chosen must also be unambiguous and familiar to all users of the catalogue. Similarly, if there are variant spellings of the same term (e.g. pediatric/paediatric) or different possible forms of the same heading (e.g. air pollution/ pollution/pollution of air), only one is used as the heading. If several meanings are attached to



one term, that term must be qualified so that it will be clear to the users for which the meaning is intended. As for example, Crane (Bird)/Crane (Lifting equipment).

### **Consistent and Current Terminology**

It follows from what has been said regarding the justifications for uniform headings that the terminology in headings should be both consistent and current. Two problems are particularly important here: choice among synonyms and change in usage.

By principle, common usage prevails when there is a problem of choices among synonymous terms. Standard lists of subject headings designed for general collections prefer a popular term rather than scientific one. In such a situation, a research or special library having specialized library collection and clientele can make extensive modifications of standard lists. Changes in usage also present many practical difficulties. A term chosen on the basis of common usage may become obsolete with the passage of time. Subsequently, a list of subject headings may incorporate current terminology as long as entries pose a problem because of the large number of entries listed under the existing subject headings. In such a situation a subject authority file is to be maintained. Once a heading is changed, every record that was linked to the old heading can be linked to the new heading and this decision is recorded in the subject authority file.

### **Form Heading**

In addition to the subject headings, there are form headings that have the same appearance as topical subject headings but refer to the literary or artistic form (e.g., Essays, Poetry, Fiction, etc.). Libraries that want to provide access to these kinds of materials may assign appropriate form headings to individual works as well as to collections and materials about the form. Apart from literary works themselves there are also many kinds of library materials about literary forms that require subject headings. For a document on how to write an essay, the heading "Essay" represents a subject and subject headings and form headings can be made by using the singular form for the topical subject heading and plural for the form heading (e.g., Short story, Short stories). In addition to the literary form headings, there are some other form headings that are determined by the general format and purpose of the documents, such as Almanacs, Encyclopaedias, Dictionaries, and Gazetteers.

### **Cross Reference**

Cross-references direct the user from terms not used as headings to the term that is used, and from broader and related topics to the one chosen to represent a given subject. Three types of cross-references are used in the subject headings structure. These are discussed below:

#### **a) See (or USE) references**

These references guide users from terms that are not used as headings to the authorized headings for the subject in question. 'See' or 'USE' references ensure that in spite of different names for (or different forms of the name of) a given subject a user shall still be able to locate materials on it.

#### **b) See also (including BT, NT, and RT) references**

These references guide users to the headings that are related either hierarchically or associatively and are used as entries in the catalogue. By connecting related headings, the 'see also' (RT, for related term) references draw the user's attention to material related to his interest. By linking hierarchically related headings, 'see also' (BT, for broader term; NT, for narrower term) references help the user to search specific deviations or aspects of his subject of interest.

#### **c) General references**

General references direct the user to a group or category of headings instead of individual headings. It is sometimes called a 'blanket reference'. The provision of general references in the standard list of subject headings obviate the need to make long lists of specific references and thus ensure economy of space.

3) State the principles that guide the indexers in the choice and rendering of subject headings from the Standard List.

ii) Check your answers with the answers given at the end of this Unit.

This image shows a full page of primary-ruled paper. It features multiple sets of horizontal lines designed to help children learn letter height and placement. Each set consists of a solid top blue line, a dashed middle blue line, and a solid bottom blue line. These sets are repeated down the entire page, providing ample space for practicing writing letters and words. The margins are consistent on all sides.

On local level, a library devises such a means, which ensures conformity and currency of the subject libraries following one of the standard lists of subject headings (e.g. LCSH or Sear's List). It creates local subject authority records only for headings not yet appearing in established headings in a subject heading list along with needed maintenance information. Thus, the subject authority control at the local level includes correcting erroneous headings and cross-references, updating obsolete headings, and adding or revising cross-references necessitated by new headings.

15.4.3 Functions of a Subject Authority File

The functions of a subject authority file are discussed below:

**Cataloguing:** The subject authority file serves as the source of indexing vocabulary and as the means of verifying or validating headings assigned to individual cataloguing records. It helps to ensure that a) the same heading is assigned to all works on the same subject, b) each heading represents only that particular subject, and c) all headings assigned to cataloguing records conform to the established forms.

**Maintenance:** Necessary adjustments to cataloguing records are needed to be added from time to time as a result of changes in the indexing vocabulary. When existing subject headings are revised or new headings are added; cross-references are often affected and should be adjusted. The subject authority file reflects the most current status of headings and cross-references and thus, serves as the source for verification and validation of subject headings as to the cataloguing records. It is also useful when a library converts its manual form to the online mode and wishes to have previously existing records reflect current practice.

**Retrieval:** Subject authority file helps the users in two ways. First, subject headings displayed in the subject authority file show the user the terminology and form of subject access points in the catalogue. Second, the cross-references guide the users to related headings when user's input terms fail to retrieve useful records.

Self Check Exercise

4) Mention the importance of a subject authority file to the cataloguer.

**Note:** i) Write your answers in the space given below.  
ii) Check your answers with the answers given at the end of this Unit.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

15.5 THESAURUS

The word 'thesaurus' comes from Greek term 'thesauros' meaning a storehouse or treasury of words. The Oxford English Dictionary defines "thesaurus" as a archaeological term "a treasury of temple, etc." and quotes its use in 1736 as a treasury or store house of knowledge. Dictionary defines it as "a book of words or of information about a particular field or a set of concepts, specially a dictionary of synonyms". A dictionary lists words along with their meanings; synonyms, etc. in alphabetical order, but a thesaurus assembles all words related to an idea at one place. Modern usage may be said to date from 1852 when Peter Mark Roget thought of his thesaurus as a classification of ideas. Roget's Thesaurus had nothing to do with information retrieval. But his novel idea was later profitably employed in the compilation of thesaurus for information retrieval.

Helen Brownson is said to be the first person who used the term 'thesaurus' in the context of information retrieval in a paper presented at the Dorking Conference on Classification Research in 1957. H. P Luhn was probably the first person to think in terms of information retrieval thesaurus, who suggested the compilation, for indexing purposes, of 'families of notions' and dictionary of 'notion families'. The first thesaurus used in information retrieval system was developed by Du Pont in USA around 1969 and since then many thesauri have been brought out in different subject fields: A number of standards have also come into existence to provide guidelines in the design and development of monolingual and multilingual: thesauri.

15.5.1 Definition

There are a number of definitions of 'thesaurus' provided by different experts and organisations. The most comprehensive one has been provided by the International Standards Organisation (ISO) as the basis of structure and functions of a thesaurus.





In terms of functions, it states "a thesaurus is a terminological control device used in translating from the natural language of documents, indexers or users into a more constrained 'system language' (documentation language, information language)". In terms of structure, the Standard says, "a thesaurus is a controlled and dynamic vocabulary of semantically and generically related terms which covers a specific domain of knowledge".

In short, a thesaurus may be defined as a compilation of descriptors for use in an information retrieval system arranged in an alphabetical order and manifesting the various types of relationships existing between the descriptors. An information retrieval thesaurus is a kind of semantic networking of concepts.

### 15.5.2 Functions of a Thesaurus

The major functions of a thesaurus include the following :

- a) it provides a standard vocabulary for a given subject field by exercising control on the vocabulary of terms used in an indexing language. Methods of controlling the vocabulary .are:
  - i) out of all possible synonyms and quasi-synonyms, only one term is selected as a descriptor,
  - ii) the scope of the meaning of the term is clearly indicated in a scope note for the best suitability of the selected meaning,
  - iii) a definite rule is followed for compound terms, word-forms, number (singular/plural) and spellings are standardized, and
  - iv) homonyms are differentiated by qualifiers;
- b) it shows the intrinsic, semantic relationship existing between, terms, and thus provides system of references between terms;
- c) it helps the indexer and the searcher in the choice of preferred terms;
- d) it provides hierarchical display of terms so that a search can be broadened or narrowed systematically;
- e) it increases the speed of retrieval by use of indexing terms and search terms; and
- f) it provides a map of a 'given subject field, which helps to understand the structure of the field.

### Self Check Exercises

- 5) What do you mean by "Thesaurus"?
- 6) Discuss the functions of a thesaurus.

**Note:** i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

### 15.5.3 Types of Thesauri

Based on the nature of terminology control, there are mainly two types of thesauri:

- a) controlled thesauri which allow only one term (preferred term) to denote a concept for the purpose of indexing and searching; and



- b) free language thesauri, which allow use of all terms to denote a concept to be used for indexing and searching.

The controlled thesauri can be maintained manually but free language thesauri require machine maintenance and retrieval.

#### 15.5.4 Structure of Thesaurus

The internal form of individual entries and the arrangement of various entries in relation to one another constitute the structure of a thesaurus. Cross-references make explicit the way in which entries relate to each other in a network of concepts. Each entry in a thesaurus consists of a pack of terms, which are related to it in different ways. The different terms in the entry are displayed in the following format:

DESCRIPTOR

(With scope note whenever needed)

Synonyms and quasi-synonyms

(displaying equivalence relationship and denoted by the relationship indicator USE/UF (Use For))

Broader Terms

(displaying hierarchical - subordinate relationship and denoted by BT)

Narrower Terms

(displaying hierarchical - subordinates relationship and denoted by NT)

Related Terms

(displaying associate relationship and denoted by RT)

Top Term

(displaying hierarchical - subordinates relationship and denoted by TT . Top term or TT is not repeated when all the descriptors belong to the same broad class).

A thesaurus may be either alphabetical, or classified, and it may or may not include a graphical display. In an alphabetical thesaurus, the descriptors followed by their relationships are listed in alphabetical sequences. In a classified thesaurus, the descriptors are listed in accordance with the hierarchical relationships represented in the thesaurus. The various levels of hierarchy are shown by appropriate indentations. The graphical displays are multi-dimensional ways of representing the relationships between terms. Such relationships are indicated by arrows lines or by presenting term in concentric circles showing hierarchy.

Reciprocal entries appear for each term in a thesaurus whenever a relationship, whether hierarchical or non-hierarchical, is established between two terms.

#### 15.5.5 Relationships Between Terms

The inter-relationships between the terms in a thesaurus are brought out by two basic types of relationships - (1) Hierarchical relationships, and (2) Non-hierarchical relationships.

##### Hierarchical Relationships

Hierarchical relationship refers to superordinate and subordinate relationship for a concept. This relationship may be of three types:

Genus-Species (Generic) relationship

Hierarchical Whole-Part relationship

Instance relationship

- a) **Genus - Species (Generic) Relationship** links genus and species and represents the basis of scientific, taxonomic system. As for example,

Rodents - Mice

NT Mice - BT Rodents

Here, 'Rodents' is a genus (broader concept) while 'Mice' represents its species (narrower concept).



- b) Hierarchical Whole Part Relationship** means that the name of a part implies the name of its whole in any context. As for example,

Cardio-vascular System		Circulatory System
BT Circulatory System		NT Cardio-vascular System

Here 'Cardio-vascular system' always refers to a part of its whole 'Circulatory System'.

- c) Instance Relationship** occurs in a particular instance, which links proper name with common noun.

As for example,

Mountain Regions		Alps
NT Alps		BT Mountain Regions
Mountain		Himalayas
NT Himalayas		BT Mountain Regions

Instance relationship is often not shown in a thesaurus to control its size.

### **Non-hierarchical Relationship**

When two terms are related other than hierarchical, the relationship may be called non-hierarchical relationship. This relationship may be further grouped as:

- a) Equivalence (or preferential) relationship, and
- b) Associative (or affinitive) relationship.

#### **a) Equivalence (or Preferential) Relationship**

Usually it refers to the preferred terms and distinguishes such terms from the non-preferred terms. When terms are regarded as similar or almost the same in meaning, they can be combined with the same concept. Synonyms, near-synonyms, and quasi-synonyms come under this category. The symbols used to represent these relationships in a thesaurus are USE and OF (Used For). As for examples,

Disabilities	Handicaps
USE Handicaps	OF Disabilities
Games	Sports
USE Sports	OF Games

#### **b) Associative (or Affinitive) Relationship**

Associative or affinitive relationship is very difficult to define. This relationship is employed to cover other relationship between terms that are related but are neither consistently hierarchical nor equivalent. In other words, here two terms are conceptually associated on a number of different basis while satisfying the requirement that one of the terms should function as a component in any explanation or definition of the other. They are indicated by the code RT (Related Terms). Some examples of associative relationships are demonstrated below:

- i) A discipline or field of study and the object or phenomenon studied
 

Birds	Ornithology
RT Ornithology	RT Birds
- ii) An action and its property:
 

Indexing	Efficiency
RT Efficiency	RT Indexing
- iii) An action and resulting product:
 

Weaving	Cloth
RT Cloth	RT Weaving
- iv) Coordinate ideas:
 

Classification	Cataloguing
----------------	-------------

	RT Cataloguing	RT Classification
v)	Ideas having common elements in their definition:	
	Management	Administration
	RT Administration	RT Management

**Self Check Exercises**

7) What are the different types of relationships displayed in a thesaurus?

- Note:** i) Write your answers in the space given below.  
ii) Check your answers with the answers given at the end of this Unit.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

**15.5.6 Thesauri and Subject Headings List**

Both thesauri and subject headings lists are vocabulary control devices, but they are used in different situations. The essential characteristics which differentiate them are considered in the following sections:

- a) Subject headings list fulfills the needs of pre-coordinate indexes, whereas a thesaurus is designed to meet the specific needs of post-coordinate indexes.
- b) Thesauri generally contain terms that are more specific than those found in conventional subject headings list,
- c) A thesaurus normally avoids inverted terms such as 'Psychology, Children'.
- d) The relationship display is more extensive in the case of thesaurus than in the case of traditional subject headings list. Incidentally, some well-known subject headings list, such as Library of Congress Subject Headings and Sear's List of Subject Headings, in their latest editions, have adopted thesaurus format and are, thereby showing the relationships existing between terms.
- e) The relationships between terms listed in a thesaurus are not transferred to the indexes in many cases. Dictionary catalogues normally provide 'See' and 'See also' references linking the related subject headings.

The above are some of the significant aspects that distinguish a thesaurus from a conventional subject headings list.

**15.5.7 Thesauri and Classification Schemes**

A classification scheme, especially a faceted and hierarchical one, is able to show hierarchical, faceted and phase relationships, but often misses other associative and equivalence relationships. However, the real difference between a classification scheme and a thesaurus lies in their purpose and use.

**15.5.8 Thesauro-Facet**

This concept has been developed by Jean Aitchison and others for English Electric Company. It is basically a faceted classification, integrated with a thesaurus. Thesauro-facet consists of two sections: a) faceted classification scheme, and b) alphabetical thesaurus. Here, the thesaurus replaces the alphabetical subject index, which normally follows the schedules in a conventional faceted classification. Terms appear twice - once in the schedule and once in the alphabetical

thesaurus, the link between two locations being the notation or class number. It can be used in both pre-and post-coordinate indexing systems.



### 15.5.9 Classaurus

It is also a vocabulary control device developed by Dr. Ganesh Bhattacharyya at DRTC that incorporates in itself features of both a faceted classification scheme as well as that of a conventional alphabetical thesaurus. It is an elementary category-based (faceted) systematic scheme of hierarchical classification in verbal plane incorporating all the necessary and sufficient features of a conventional information retrieval thesaurus. Like any classification scheme, it displays hierarchical relationships among terms in its schedules. Like a faceted classification scheme, there are separate schedules for each of the Elementary Categories (Entity, Property, and Action) and for common modifiers like Form, Time, Place, and Environment. Like any thesaurus, each of the terms in the hierarchic schedules is enriched by synonyms, quasi-synonyms, etc. Unlike a thesaurus, a classaurus does not include other associatively related terms (RTs) because of its category-based (faceted) structure. It is said that a term in one elementary category has a high chance of being associatively related with another term in another category depending on the subject of the document. It is assumed that RTs should not be dictated by the designer of the classaurus, rather it should be dictated by the document itself since any term may be associatively related to other terms depending on the nature of the thought content of the document. The classaurus has two parts: the Systematic Part and the Alphabetical Index part.

The Systematic Part consists of common modifiers. Each entry in the Systematic Part consists of Descriptor; Definition/Scope Note (whenever needed), Synonyms (UF), if any, parts, and species/ types. The hierarchy of terms is shown by indentation indicated by `(.dot). Terms in an array are arranged according to alphabetical order. Each term in the Systematic Part is assigned a unique alphanumeric code. The Alphabetical Index Part to the classaurus contains each and every term including synonyms occurring in the Systematic Part along with its address (i.e. alphanumeric code).

### Self Check Exercises

- 8) Explain the difference between a thesaurus and a standard list of subject headings.
- 9) Write the essential features of "Thesauro-facet" and "Classaurus".

**Note:** i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

### 15.5.10 Role of a Thesaurus in Indexing and Searching

The thesaurus contributes to consistency in indexing between indexers and reduces incorrect term assignments to a great extent because of the provision of the term definition or scope note. This *is* particularly true if descriptors, which are some what related, are carefully distinguished by scope notes.



Thesaurus plays the following two roles

- a) prescriptive role. It prescribes as to what term should be assigned. The 'use' reference is a prescriptive indicator.
- b) suggestive role. It suggests terms to be considered instead of, or in addition to etc. The RT (and to a certain extent the BT and NT) are suggestive indicators.

#### 15.5.11 How to Use a Thesaurus

The final process of indexing involves translating the relevant subject concepts of the document into the indexing language or the descriptor language of the system.

The translation stage involves the comparison of each term in the subject statement with the descriptors of the thesaurus to determine:

- a) whether an exact equivalent is available in the thesaurus,
- b) whether a broader term in the thesaurus is adequate for retrieval,
- c) whether the pre-coordinated term exists in the thesaurus (e.g. corrosion prevention, production capacity) which can be substituted for one or more terms (adjacent to each other) in a subject statement,
- d) whether it is necessary to introduce a new term as a last resort. New terms should be introduced only after a careful consideration of terms already available in the thesaurus, the relative frequency of occurrence of the concept in the literature, the relative frequency of its probable use in the system, its relationship to descriptors already existing in the thesauri and its technical precision and acceptability as determined by references to other thesauri, dictionaries, handbooks, glossaries and the like.

After each term in a subject statement is translated into the descriptors from the thesaurus used in a system, a set of references is established between the terms in a subject statement and the corresponding terms revealed by the thesaurus. This step is essential when the thesaurus is used as a vocabulary control tool in a system which adopts as pre-coordinated indexing system such as, PRECIS, POPSI or BTI. The two kinds of references are:

- a) *See* reference (corresponding to USE reference) which directs the user to an appropriate term from which is not used as a descriptor. A *see* reference is made for the non-preferred synonyms or quasi-synonymous term to the preferred lead term.
- b) *See also* reference (corresponding to BT to NT) to the lead term is made from its immediate broader term. A *see also* reference is made from the other terms which are associatively related (M) terms, having strong mental association between that term and the lead term.

#### 15.5.12 Thesaurus in Post-Coordinate Indexing Systems

Thesaurus has come into existence as a tool for vocabulary control in post-coordinated indexing system. It helps the indexer in selecting the appropriate terms for indexing, as the indexer may not be familiar with the subject field. It also aids the indexer in his choice of search terms, especially in searching a field with which he is not intimately familiar.

Besides providing alternative search paths, a thesaurus aids in choosing terms for summation. When information on all the aspects of a subject field is required, search has to be made covering all the related aspects. Here the searcher may not be able to recall all the information that the system can give without the aid of a thesaurus.

In the case of computer searching, the links between words used in documents and words used as descriptors in indexing are assembled via a thesaurus.

---

### 15.6 SUMMARY

In this Unit we have dealt with the vocabulary control with particular reference to the Subject Headings List and Thesauri. The meaning of vocabulary control, its need, objectives and important requirements have been discussed. The subject headings list has been explained in the context of the general principles associated with the choice and rendering of subject headings. Subject authority file has also been considered in the light of use and maintenance of thesauri, their types, functions, structure, nature of relationships, between terms, and the important role they play in indexing and searching have also been discussed.

## 15.7 ANSWERS TO SELF CHECK EXERCISES



- 1) Vocabulary Control means selecting terms from a set of terms which are used in an information retrieval system. Here, a controlled set of terms is selected from a natural language and used to represent the subject matter of documents and requests of users at the time of indexing and searching respectively. A complete one to one relationship between the concepts and terms is established by controlling the vocabulary. Synonyms, homonyms, word form, spelling, etc. are controlled in order to standardize the terminology to be used - for indexing and searching. A certain degree of structure is introduced in a controlled vocabulary so that terms whose meanings are related are brought together or linked in some way. Out of the synonyms, only one term is accepted as a standard term (i.e. descriptor or preferred term) and this is used in subject description. References are made from non-preferred terms to preferred terms facilitating search by the users. Controlled vocabulary can be applied in any situation in which the standardization of terminology is needed.
- 2) The objectives of vocabulary control can be summarized as follows : first, to ensure the consistent representation of the subject matter of documents at the time of indexing and at the time of searching. It is achieved by controlled synonymous and nearly synonymous expressions and by distinguishing among homographs. Second, to facilitate the conduct of a comprehensive search on some topic by linking together in some way the terms that are most closely related semantically.
- 3) The principles that guide the indexers in the choice and rendering of subject headings from the Standard List are:
  - a) **Specificity:** A work (document) should be entered under the most specific subject heading that accurately and precisely represents its content.
  - b) **Common Usage:** The heading chosen to express a subject must represent common usage of the class of readers for whom the material on the subject within which the heading falls is intended.
  - c) **Uniformity:** One uniform heading must be selected from several synonyms, and this heading must be applied consistently to all works on the topic.
  - d) **Form heading:** Form heading should be chosen for the work representing its physical, bibliographic, artistic or literary form and for the work about literary form that requires subject heading.
  - e) **References:** Adequate references should be made to direct the user from unused heading to used headings, and from headings referring broader and related topics to the headings chosen to represent a given subject.
- 4) Subject authority file is an important means of maintaining control over the vocabulary and references. It serves as the source of indexing vocabulary and as the means of verifying or validating heading assigned to individual cataloguing records. By consulting the subject authority file, the cataloguer can make necessary adjustments when existing headings are revised or new headings are added. The subject authority file is particularly useful in ensuring uniformity and consistency in subject heading terminology and cross-references.
- 5) Thesauri is a vocabulary control device used in selecting standardized term(s) for representing the subject matter of a document at the time of indexing and for representing the request at the time of searching. It is a list of terms arranged in some meaningful form and provides hierarchically and non-hierarchically (Equivalence and Associative) related terms for a specific domain of knowledge:
- 6) The functions of a thesaurus may be summarized as follows:
  - i) To provide standard vocabulary for a given subject area by exercising control on the vocabulary of terms used in an indexing language:
  - ii) To provide an aid to indexers in selecting term(s) for describing the subject matter of documents.
  - iii) To help users to formulate their queries precisely.'
  - iv) To provide display of-relationships between terms to facilitate the conduct of a comprehensive search systematically.
  - v) To provide a system of references between terms which will ensure that only one term from a set of synonyms is to be used both for indexing and searching.



- 7) Two types of relationships are displayed in a thesaurus: hierarchical and non-hierarchical. Hierarchical relationship indicates the interrelationship between concepts in a hierarchy. It expresses degrees or levels of super ordination and subordination of concepts. This relationship may be of three types: Genus-species, hierarchical whole-part and instance relationships. Non-hierarchical relationship may be further grouped as - Equivalence relationship and Associative relationship.
- 8) Both thesauri and subject headings lists are vocabulary control devices utilised in information retrieval. But, they are utilised in different environments.
  - i) thesauri contain terms that are more specific than those found in subject headings lists.
  - ii) thesauri fulfill the needs of post-coordinate indexes, whereas a subject headings list fulfils the needs of pre-coordinate indexes.
  - iii) thesauri tend to avoid inverted terms such as 'Art, French'.
  - iv) Headings in thesauri are not subdivided. But, headings of this type are very common in conventional subject headings lists.
  - v) The relationship display in a thesaurus is more extensive than the relationship display in a subject headings list.
  - vi) The relationships between the terms listed in a thesaurus are not in many instances transferred into the index where as alphabetical subject catalogues usually contain 'See' and 'See also' references\_
- 9) Both 'Thesauro-facet' and 'Classaurus' are the extended form of a thesaurus. Thesauro-facet consists of. i) a faceted classification scheme which displays the generic' relations between them and from which compound subjects can be synthesized; and ii) an alphabetical thesaurus. The link between two parts being the class number. It can be used in both pre-and-post-coordinate indexing systems. 'Classaurus' incorporates in itself features of both a faceted classification scheme as well as that of a conventional thesaurus. A classaurus does not display associatively related terms. The classaurus consists of two parts: i) the systematic part consisting of schedules of different elementary categories and common modifiers; and ii) an alphabetic index part containing each term including synonyms occurring in the systematic part. The link between the two parts being the alphanumeric code assigned to each term in the systematic part.

## 15.8 KEY WORDS

<b>Associative Relationship</b>	:	Relationships between pairs of terms which are not member of an equivalence set, nor can they be organized as a hierarchy in which one term is subordinated to another. Yet, they are mentally associated to such an extent that the link between them is made explicit in the thesaurus by means of relationship indicator 'RT' (Related Term).
<b>Authority Control</b>	:	The process of maintaining consistency in access points in a catalogue.
<b>Broader Term</b>	:	The term denoting super-ordinate idea in a hierarchy.
<b>Classaurus</b>	:	An elementary category-based faceted systematic scheme of hierarchical classification in the verbal plane having all the necessary and sufficient features of a conventional alphabetical, thesaurus excluding the enumeration of the Related terms (RT).
<b>Cross-reference</b>	:	A direction from one heading or entry to another.
<b>Descriptor</b>	:	See Preferred Term.
<b>Direct Entry</b>	:	Entry of a document directly under the most specific term (i.e. subject heading) that accurately and precisely represents its content.
<b>Equivalence Relationship</b>	:	This is the relationship between preferred and non-preferred terms, where two or more terms are





	regarded, for indexing purposes, as referring to the same concept.
<b>Form Heading</b>	: A heading representing the physical, bibliographic, artistic, or literary form of a work, e.g. encyclopaedias and dictionaries, essays, short stories, etc.
<b>General Reference</b>	: A blanket reference to a group of headings rather than a particular heading.
<b>Hierarchical Relationship</b>	: Relationship between terms based on the super-ordination and subordination of concepts, where the super-ordinate term represents a class or a whole, and subordinate terms refer to its members or parts.
<b>Literary Warrant</b>	: The principle which allows the use of an actual collection or holdings of a library or actual published work as the basis of or for developing a classification scheme or thesaurus.
<b>Narrower Term (NT)</b>	: The term denoting sub-ordinate idea in a hierarchy.
<b>Non-hierarchical Relationship</b>	: Relationship between terms not based on super-ordination or sub-ordination of concepts, but refers to either associative or equivalence relationship.
<b>Non-preferred Term</b>	: Synonyms or quasi-synonyms of terms which are not assigned to documents, but which are provided as entry points in a thesaurus.
<b>Preferred Term</b>	: Term used consistently to represent a concept in indexing; sometimes known as 'descriptor'.
<b>Related Term (RT)</b>	: The term referring mental association to other term while satisfying the requirement that one of the terms is a necessary component in any explanation or definition of the other, to the extent that the term 'Birds', for example, forms a necessary part of the explanation of 'Ornithology'.
<b>Specific Reference</b>	: A reference from the one heading to another.
<b>Subject Authority File</b>	: A collection of subject authority records.
<b>Subject Authority Record</b>	: A record of a subject heading that shows its established form, cites the authorities consulted in determining the choice and form of the heading, and indicates the cross-references made to and from the heading.
<b>Subject Headings List</b>	: A list of subject headings or terms, including references, to be used as a standard cataloguing or indexing.
<b>Thesauro-Facet</b>	: Integration of faceted classification schedule and thesaurus. The thesaurus serves as an index to the classification schedule.
<b>Thesaurus</b>	: It is a compilation of words and phrases showing synonyms, hierarchical and other relationships and dependencies, the function of which is to provide a standard vocabulary of information storage and retrieval systems. The object of the thesaurus is to exert terminology control in indexing and to aid in searching by alerting the searcher to the index terms that have been applied.
<b>Top Term (TT)</b>	: Term referring to the broadest class or top most super-ordinate idea in a hierarchy.
<b>User Warrant</b>	: The principle, which allows the use of users' requests for information as the basis for developing a classification scheme or thesaurus.
<b>Vocabulary Control</b>	: A controlled set of terms used to represent the subjects of documents in indexing, and to search for these documents in a particular system.



---

## 15.9 REFERENCES AND FURTHER READING

---

- Aitchison, J. (1970). The Thesaurifacet: A Multipurpose Retrieval Language Tool. *Journal of Documentation*. 26; 187-203
- Aitchison, J. and Gilchrist, A. (1987). *Thesaurus Construction: A Practical Manual*. 2nd ed. London : ASLIB.
- Bhattacharyya, G. (1982). *Classaurus: Its Fundamental, Design and Use*. (Paper presented to the Fourth International Study Conference on Classification Research. Augusburg, June 28-July 2, 1982).
- Chakravorty, A.R. and Chakraborty, Bhubaneswar. (1984) *Indexing: Principles, Processes and Products*. Calcutta : World Press.
- Chan, Lois Mai (1994). *Cataloguing and Classification: An Introduction*. 2nd ed. New York : McGraw-Hill.
- Clack, D.H. (1990). *Authority Control: Principles, Applications and Instructions*. Chicago : American Library Association.
- Foskett, A.C. (1996). *The Subject Approach to Information*. 5th ed. London : Library Association Publishing.
- Foskett, D.J. (1979). Thesaurus. In: Kent, Allen [et al.] (eds.) *Encyclopaedia of Library and Information Science*. Vol. 25.
- Ghosh S.B, and Satparthi, J.N. (ed.) (1998). *Subject Indexing Systems*. Calcutta: IASLIC
- Lancaster, F. W. (1985). *Vocabulary Control for Information Retrieval*. 2nd ed. Arlington, Va. : Information Resources Press.
- Soergel, D. (1974). *Indexing Language and Thesauri: Construction and Maintenance*. Los Angeles, California Melville Publishing.