Efficiency and effectiveness of deep structure based subject indexing languages: PRECIS vs. DSIS

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/7280

Publisher: © Subal Chandra Biswas

Please cite the published version.
This item is held in Loughborough University’s Institutional Repository (https://dspace.lboro.ac.uk/) and was harvested from the British Library’s EThOS service (http://www.ethos.bl.uk/). It is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/

Disclaimer:
EFFICIENCY AND EFFECTIVENESS OF DEEP STRUCTURE BASED SUBJECT INDEXING

LANGUAGES: PRECIS VS. DSIS

by

Subal Chandra Biswas, B.A.(Hons.), M.A., M.L.I.S.

A doctoral thesis

submitted in partial fulfilment of the requirements

for the award of the degree of

Doctor of Philosophy

of the

Loughborough University of Technology

December, 1988

Supervisor of Research: Frederick Smith, B.A., F.L.A.

Department of Library and Information Studies

Loughborough University of Technology

© Subal Chandra Biswas, 1988
Dedicated to my parents

and my wife, Lalita
ACKNOWLEDGEMENTS

I would like to thank Mr F. Smith for his supervision and collaboration (in the form of developing all the software necessary for this project) in the present study. He also deserves special thanks for the pain he took in producing this thesis in its present format.

Professors P. Havard-Williams and A. J. Meadows were respectively my Directors of Research, and I am grateful for their direction, encouragement and provision of facilities for completing this research.

I received help and encouragement from a number of persons within the DLIS, and in particular from Mrs I. A. Smith, Dr J. D. White and Mrs G. Tseng.

I am thankful to Dr F. J. Devadason, Assistant Director (Information), Central Leather Research Institute, Madras, India (formerly of Documentation Research and Training Centre, Bangalore, India), for his replies to all my queries regarding DSIS and Classaurus, and for his endurance and encouragement during my meeting with him in Madras during March 1987.

Financial support for the present work was provided by a scholarship under the Commonwealth Scholarship Scheme offered by the Commonwealth Scholarship Commission in the United Kingdom. The debt is gratefully acknowledged. I am also indebted to my employer, the University of Burdwan, Burdwan (W. B.), India, for providing the leave of absence during the period of research.

I must also acknowledge my intellectual debt to my ex-teacher, Dr M. Mahapatra, Utkal University (formerly of University of Burdwan), India, from whom I took my first lessons in research.

Finally, I must thank, for the encouragement I have received, all my
family, friends and relations, both here and in India, and in particular to my parents (who always carried the burden of educating me), and to my wife, Lalita (for her understanding and patience).
ABSTRACT

A 'Subject Indexing Language' (SIL) is an artificial language used for formulating names of subjects. Although classificationists have sought for universals in many fields of study such as, philosophy, biology, general systems theory, etc., the search for a deep structure of SILs formally began with Ranganathan's idea of 'absolute syntax' and was brought to the present by G. Bhattacharyya and D. Austin. Whereas Bhattacharyya's deep structure of SIL is primarily based on classificatory principles (parallel to 'absolute syntax'), the deep structure proposed by Austin has a linguistic connotation.

The present study describes and compares two such deep structure-based SILs, viz., PRECIS (PREserved Context Index System) and DSIS (Deep Structure Indexing System), a recent computerized version of POPS (POstulate-based Permutated Subject Indexing), developed by F. J. Devadason at Documentation Research and Training Centre, Bangalore, India. Both also belong to the category of SILs typified as 'string indexing' languages. The study involves: i) writing of a suitable DSIS index entry generation program, ii) using both PRECIS (in-house) and DSIS programs to index a collection of representative sample documents from the soft sciences, iii) analyzing and comparing their respective syntactic and semantic aspects in terms of both linguistic and classificatory principles, and iv) applying some measures of efficiency and effectiveness. It was realized that certain modifications in the existing DSIS string manipulation algorithms are necessary to make the program fully operational. Although, no attempts have been made to quantify the measures of effectiveness and efficiency as such, suggestions have been provided as to what these probably would be. Some indications of their searching difficulties for a prospective searcher have been put forward as well.
Keywords

Index language universals
Subject indexing languages
PRECIS
Deep Structure Indexing System
Classed vocabularies

Information languages
String indexing languages
POPSI
Controlled vocabularies
Classaurus
CONTENTS

Declaration iii
Acknowledgements iv
Abstract vi
List of Figures xiii
List of Appendices xiv

Chapter 1: Introduction 1

Chapter 2: Subject Indexing Language 7

2.1 The Human Communication Process 7

2.2 Information Language (IL) 9

2.2.1 Information Language vs. Natural Language 10

2.3 Universals in Information Languages 13

2.3.1 Linguistic Universals in Information Languages 15

2.3.1.1 Evidence from Theoretical Linguistics 19

2.3.1.2 Cognitive Aspects of Language Processing 22

2.3.2 Evidence from Related Disciplines 28

2.3.3 Universal Forms and Subject Representation 30

2.3.3.1 Ranganathan and the Indian School 31

2.3.3.2 Classification Research Group (UK) 34

2.3.3.3 Comparative Method of Finding Universals 37

2.3.3.4 Intermediary IL or Switching Language 38

2.4 A Note on String Indexing 40

Chapter 3: An Introduction to PRECIS 52

3.1 Background and Objective 52

3.2 Nature of and Relationship between Concepts 55

3.3 Syntax 57
3.3.1 Entry format

3.3.1.1 Standard Entry Format

3.3.2 Compound Terms -- Technique of Differencing

3.3.2.1 Preceding Differences

3.3.2.2 Following Differences

3.3.2.3 Time as a Difference

3.3.2.4 Universality of Differences

3.3.3 Schema of Operators

3.3.3.1 Contributions of Transformational-Generative Grammar

3.3.3.2 Evidence from Case Grammar

3.3.3.3 Deep Cases and Role Operators

3.3.4 Special Routines with Linguistic Orientation

3.3.5 Stages of Concept Analysis, and the Use of Operators

3.3.5.1 Dependent Elements

3.3.6 Inverted Format: Extra-core Operators

3.3.7 Entries Generated by the Predicate Transformation

3.3.7.1 Role Defining Terms and Directional Properties

3.3.7.2 Author-attributed Associations

3.3.7.3 Two-way Interactions

3.4 Need for Alternative Explanations

3.4.1 Time as an Organizing Principle

3.4.2 The Primitive Relationships

3.5 Semantics

Chapter 4: Introduction to DSIS

4.1 Subject Indexing Language

4.2 Deep Structure of Subject indexing Languages

4.2.1 Elementary Categories
4.2.2 Subdivisions of Manifestations
   4.2.2.1 Modifier: Compound and Complex Term
   4.2.2.2 Composite Term
4.2.3 Syntax of DSIS
4.2.4 Indicators of Deep Structure
4.3 Formulation of Name of Subject
   4.3.1 Concept Analysis and Formalisation
   4.3.2 Modulation and Standardization
   4.3.3 Category Indication
4.4 Subject Index Entry
   4.4.1 Functions of Subject Index Entry
   4.4.2 Structure of Subject Indexing Entry
   4.4.3 Cross Reference Entry
   4.4.4 Permuted Cross Reference Entry
4.5 Formation of Subject Headings
   4.5.1 Selection of Lead Term
   4.5.2 Selection of Context Term
   4.5.3 Upper Link Specifiers to Lead Term
   4.5.4 Default Lead and Context
   4.5.5 Processing Codes
4.6 Coding of Name of Subjects
   4.6.1 Coding Conventions
   4.6.2 Coding Illustration
4.7 Types of Indexes
4.8 Vocabulary Control
   4.8.1 Construction of Classaurus
4.9 Salient Features of DSIS

Chapter 5: Comparative Study

5.1 Index Language Comparisons
Section 5.2: Comparisons between PRECIS and POPSI

Section 5.3: The Present Comparative Study
  5.3.1 Objectives in General
  5.3.2 Materials and Input Preparation

Chapter 6: Subject Heading Structuring
  6.1 Handling of Compound Terms
    6.1.1 Adjectival Compounds
    6.1.2 Prepositional Phrases
    6.1.3 Complex Subjects
  6.2 Syntax
    6.2.1 Main Class or Discipline as the First Category
    6.2.2 Coordinate Relationship
    6.2.3 Viewpoint and Target
  6.3 Entry Format
    6.3.1 Choice of Lead Term
    6.3.2 Selection and Rendering of the Context Term
    6.3.3 Types of Entries
    6.3.4 Cross Reference Entries
    6.3.5 Display of Index Entries

Chapter 7: Vocabulary Control
  7.1 Alphabetical vs. Classified Structure of a Controlled Vocabulary
  7.2 Facet Analysis and Vocabulary Control
  7.3 Vocabulary Control in PRECIS and DSIS
    7.3.1 Term Structure
      7.3.1.1 Logicality of DSIS' Term Structure
      7.3.1.2 Implications of Term Structure of DSIS vis-a-vis PRECIS
    7.3.2 Representation of Basic Thesaural Relations
7.3.2.1 Hierarchical Relations 255
7.3.2.2 Associative Relationships 261
7.3.3 Categorical Structure 261
7.3.4 Need for Proper Structure in the Systematic Display 265
7.3.5 Need for Special Codes in DSIS 267
  7.3.5.1 To Indicate Coordinate Concepts 267
  7.3.5.2 To Indicate Same Parts of a Term and Its Species 268
7.3.6 Potential as a Multi-disciplinary Vocabulary Control Tool 269

Chapter 8: Searching 282
  8.1 Terminology 285
    8.1.1 Form of Terms 285
    8.1.2 Access Terms 287
    8.1.3 Index Term Context 288
    8.1.4 Compound Term as Lead And Context 290
  8.2 Single-entry vs. Multiple-entry Citation Order 292
  8.3 Some Other Aspects of Index Strings 298

Chapter 9: Summary and Conclusions 305

Bibliography 311

Exhibits 335
## LIST OF FIGURES

### Chapter 2:
- **Figure 2.1**: Model of Human Communication Process  
  - Page No: 8
- **2.2**: Case Relationships  
  - Page No: 22

### Chapter 3:
- **Figure 3.1**: PRECIS Entry Format  
  - Page No: 59
- **3.2**: Order of Terms before Shunting  
  - Page No: 60
- **3.3**: Levels of Preceding Differences  
  - Page No: 64
- **3.4**: Level of Following Difference  
  - Page No: 65
- **3.5**: Word Order in English Adjectival Compounds (Type I)  
  - Page No: 70
- **3.6**: Word Order in English Adjectival Compounds (Type II)  
  - Page No: 70
- **3.7**: Active-Passive Mood Correspondence in English  
  - Page No: 73
- **3.8**: Structure of Coordinate Relationship  
  - Page No: 87
- **3.9**: Display of the Primitive Relationships (String)  
  - Page No: 101
- **3.10**: Display of the Primitive Relationships (Entry)  
  - Page No: 102

### Chapter 4:
- **Figure 4.1**: Deep Structure of SIL  
  - Page No: 114
- **4.2**: Syntactic Structure of DS of SIL  
  - Page No: 122
- **4.3**: Structure of Subject Index Entry (DSIS)  
  - Page No: 130
- **4.4**: Chain and Links in a Name of Subject  
  - Page No: 137

### Chapter 6:
- **Figure 6.1**: Treatment of Adjectival Compounds (Type I): PRECIS/DSIS  
  - Page No: 185
- **6.2**: Treatment of Adjectival Compounds (Type II): PRECIS  
  - Page No: 186
- **6.3**: Treatment of Adjectival Compounds (Type II): DSIS  
  - Page No: 186
- **6.4**: Treatment of Prepositional Phrases: PRECIS/DSIS  
  - Page No: 189
- **6.5**: Levels of Coordination  
  - Page No: 203
Chapter 7:

Figure 7.1: Levels of Modification 250

7.2: Polyhierarchical Relationships 259
7.3: Relation between Overlapping Hierarchies 259

LIST OF APPENDICES

Appendix Page No

3.1 Schema of Operators (PRECIS) 110

4.1 Uni Component Term Lead Heading with Full Context Heading 157
4.2 Uni Component Term Lead Heading with Short Context Heading 158
4.3 Lead Heading with Upper Link Specifiers and Full Context Heading 159
4.4 Classaurus of Library Science Terms 160
CHAPTER 1
INTRODUCTION

It has been said that, the emergence of a new idea or beginning of a new tradition takes place only when some sort of conflict or contradiction is evidenced within the existing system or environment. Nicolaus Copernicus' correct description of our solar system, Karl Marx and Frederich Engels' exposition of the theory of dialectical materialism, Ranganathan's enunciation of the theory of analytico-synthetic classification, and many other similar landmarks in the history of mankind can be attributed to this very factor. It was Copernicus who first correctly described that it is the sun which holds the central position relative to earth and other planets. This was in total contradiction of the then prevalent 'geo-centric' notion of the Ptolemaic system, for contradicting which Galileo was kept under house arrest for final eight years of his life. Similarly, Marx and Engels tried to explain the progress of civilization through the process of dialectics, a theory which believes 'all things contain contradictory sides or aspects, whose tension or conflict is the driving force of change' and saw its heyday in the overthrow of the orthodox Soviet society. For S. R. Ranganathan, library classification emerged as the twin processes of analysis and synthesis, respectively, which the then prevalent systems (such as Dewey Decimal Classification, etc.) failed to show clearly. The outcome was one of the most significant developments in the field of knowledge organization, viz., the application of explicit principles and postulates, rather than some arbitrary criteria, for ordering documents and document surrogates in repositories of recorded knowledge. The origin of the piece of research described in this thesis may be attributed to a similar conjecture (certainly not of the same proportion).

The new genre of classification was very aptly named by Ranganathan as
the 'analytico-synthetic' classification, which used the technique of 'facet analysis' (also, explicitly named by Ranganathan) to achieve its end. The procedure took its first shape in the form of the fundamental categories in the schedules of his Colon Classification scheme and later in deriving the chains in his Chain Procedure (presently called Chain Indexing). Both the schemes had profound effect on the development of knowledge ordering systems (both notational and verbal) developed in Britain and Europe. During the first two decades of the second half of this century, a host of faceted classification schemes related to special subject fields were constructed. Also, Chain procedure found its way as a useful retrieval tool in large academic and commercial organizations. But, as early as the beginning of the sixties, things began to change rapidly and it was realized that the then existing theories of information retrieval were inadequate to produce a retrieval tool which can effectively meet the needs of the ordinary user as well as the specialist. Computers and automation of all methods of information handling were sought as the key to the future development of knowledge organization and retrieval. Chain procedure was found to be incompatible with mechanization. Large organizations such as the British Library began to look for better methods of bibliographic data storage, retrieval and exchange. Computers were given the responsibility of handling the age-old, tiresome, manual activities which used to constitute the major share of an information worker's day-to-day work. It was also realized that computers can take off some of the burden from the shoulder of the persons involved in non-routine work such as production of indexes and catalogues. Enhanced computer power in the form of cheap storage and fast processing capability led to another concurrent development in the field of subject indexing. Sophisticated computer software were used to generate multiple index strings from minimum of input data. This latest family of index generation systems are now generally described as 'string indexing
languages'. The development of subject indexing systems like PRECIS (PREserved Context Index System) was a direct consequence of the above trend, which *British National Bibliography* (BNB) adopted as a replacement for chain indexing. Apart from the efforts made in certain quarters to keep chain indexing still usable as a retrieval tool (e.g., in *British Technology Index* through the ingenuity of E. J. Coates), Ranganathan and his disciples (especially, A. Neelameghan, G. Bhattacharyya and, most recently, Francis Devadason) at Documentation Research and Training Centre, Bangalore, India, kept the hope aloft through further research on verbal subject indexing using facet analysis still as its foundation. The name POPSI (POstulate based Permutated Subject Indexing) saw daylight as a result of this sustained effort. Most recent developments of POPSI have been reported as part of a more versatile subject indexing system known as Deep Structure Indexing System (DSIS), which tries to assimilate together the needs of mechanization and verbal subject indexing based on facet analysis.

As Austin has observed that "On the grounds that developments in any field, including documentation, frequently represent some kind of reaction against existing systems" and took the deficiencies of chain indexing (practiced in BNB since 1950) as his point of contention in favour of PRECIS (Austin, 1974, p. 53), the origin of the present study is a similar contention expressed by Dahlberg (1986) in her foreword to a recent FID/CR report entitled *Computerized Deep Structure Indexing System*, representing the findings of Francis Devadason's (1984) solution to the problem of subject approach to information. The question was, whether Devadason's solution can or cannot, in future, replace PRECIS. However, one might ask 'why PRECIS versus DSIS?' Apart from both being string indexing systems and strongly dependent on the availability of computers, the following reasons may also be put forward as partial answers: 1) both originated during later part of the sixties, 2) both developed as a result of classification
research, hence, have strong classificatory background, 3) the first step to their origin was in the acceptance of the limitations of chain indexing, and 4) both claim to be based on some sort of universal structures.

In their recent research, the proponents of both the systems, have claimed to discover 'deep structures' (universal forms) leading to a general theory of subject indexing languages (SILs), albeit based on different principles. Austin (1982) discovered a significant parallelism between the roles used in PRECIS and systems of deep cases proposed by linguists such as Chomsky, Fillmore, etc, whereas Bhattacharyya (1980) arrived at a 'deep structure of SILs' by logically abstracting the structures of SILs of Cutter, Kaiser, Dewey and Ranganathan, which in essence parallel Ranganathan's (1967) idea of the 'absolute syntax' for subject descriptions, a sequence in which the ideas arrange themselves in the minds of majority of persons regardless of which ordinary human language they speak. Therefore, it will be appropriate if we start our discussion of the ways in which human beings communicate (both at conceptual and linguistic levels), the place of artificial languages (such as information languages) in the communication of information, and various attempts to find universal traits of information communication and dissemination (in linguistics, psychology, information science, etc.), respectively (Chap. 2). The discussion ends with a brief note on string indexing, to which both PRECIS and DSIS (including its predecessors chain indexing and POPS1) belong.

The next two chapters (3 and 4) give an overview of the two systems, PRECIS and DSIS, respectively, and can be considered rather descriptive in their nature. However, the description of PRECIS has an overtone of linguistic analysis. This has been done purposefully, because the reader can get an alternative straightforward description from a number of published sources. On the other hand, in an attempt to mitigate criticisms such as, POPS1 and its various developments were not properly documented, the chapter
The rest of the chapters describe the findings of this comparative study. Beginning with some overview of previous index language evaluation tests carried out, chapter 5 describes the various premises on which the present comparative study is carried out, and the procedures followed thereof. Chapter 6 is basically devoted to the comparison of the various syntactical aspects of both the systems, concentrating on the handling of compound terms, rules of grammar, format and display of index entries. Chapter 7 begins with an overview of controlled vocabularies using some form of classificatory principles as their base, followed by a comparison of the features generated from the respective structures of vocabulary control tools used by PRECIS and DSIS. Searching of index entries has been considered by many as the most important of the whole range of retrieval activities. Thus, some attempts have been made in chapter 8 to have an idea of the difficulties of searching the two indexes, that a prospective searcher may face. The outcome of the whole study has been briefly summarized in the final chapter (Chap. 9) along with the general conclusions which the present author has thought appropriate. The thesis ends with the bibliography and certain end matters appended in the section entitled 'EXHIBITS'.
References


CHAPTER 2
SUBJECT INDEXING LANGUAGE

2.1 The Human Communication Process

Information science is concerned with all aspects of the communication of information, language is the primary medium for the communication of information, and linguistics is the study of language as a system for communicating information (Montgomery, 1972, p. 195). Thus, one might assume that there is an interrelationship between information science and linguistics. Consequently, attempts were made to explore the relevance of linguistics to information science in general (Sparck Jones and Kay, 1973; Walker et al, 1977), and document/content analysis in particular (Gardin, 1973). However, this pre-supposition misses another aspect of the (human) communication process, i.e., the capacity of brain to produce and comprehend concepts, which is as important as its expression into language. The understanding of what can be called the process of 'concept formation', is necessary for both recording and retrieving of information, as practiced by information scientists and users of information systems, respectively. This brings us into another arena of knowledge, viz., psychology. Therefore, we can say that, the human communication process involves language, which is the spectrum from thought through speech, encompassing the auditory, visual, tactile, and gestural modes. A simplified modelling of the process is presented by Bivins (1977) (Figure 2.1).

Level 1 is the level of cognition or thought, the subject of much concern to, among others, Piaget, Bruner, and Vygotsky. Level 2 has been the traditional concern of theoretical linguists, such as Chomsky, Fillmore, etc. 'Language' as it is usually known is located at this level. We are mainly concerned with these two levels, since, the transduction from Level 1
to Level 2 in the model is the process involved in concept formation. Level 3 is an observable phenomenon, much studied by speech and hearing specialists among others, and is beyond our concern. There is an observable similarity between the three levels in the model and the three planes of classification -- the idea plane, the verbal plane and the notational plane -- as enunciated by Ranganathan (1967b, p. 327). He was correct in noting that, among the three, the idea plane is paramount. He regarded the three planes as interdependent, but failed to see that, in order to use 'ideas' or concepts as the basis for classification, one must know the process of concept formation. The same is true of the work by theoretical linguists. Ranganathan also considered that "... generally speaking, work on the verbal plane has been neglected" by traditional classificationists. So, it is clear that, there is a considerable amount of interdependence between thought and language, and this phenomenon has been largely ignored by researchers through the first half of this century. Recent work in experimental psychology has begun to demonstrate what actually may be taking place at Level 2, and at the transduction from Level 1 and Level 2.

Figure 2.1: Model of Human Communication Process

Languages are means of communication between their speakers (writers)
and hearers (readers). A natural language (NL) is one which people use in their day-to-day communication -- both spoken and written. Language as a system has its own ingredients. One of the essential ingredients of natural language is the elements or vocabulary, which comprises of a sign or combination of signs (physical sounds or written marks) used to designate the physical objects or mental constructs they refer to (referents). These elements are related to each other on two basic axes, viz., paradigmatic (semantic) axis and syntagmatic (syntactic) axis. The semantic relationships are 'fixed' by the structure of the language-system, whereas the syntactic ones are 'selected' during the utterance of particular messages -- the former are *a priori* while the latter are *a posteriori*.

### 2.2 Information Languages

Information systems are concerned with the communication of information about documents to the potential users (readers) of those documents. The means of communication are the 'information languages' (ILs) of information systems, which is synonymous to 'information retrieval language', 'information description language', 'documentary language', etc. (Soergel, 1967, p. 219; Stokolova, 1977, p. 228). An information language has been defined as any set of symbols used for expressing certain characteristics of data on which data processing is performed, in whatever field (Cardin, 1966). The data are frequently verbal data and, in this case, what is envisaged is the data from the point of view of meaning or content. Generally these languages are organized on the semantic plane, i.e. they are presented in the form of a classification system. Furthermore, these languages are often accompanied by rules which determine the permitted syntactic links, or, in other words, they possess some form of organization on the syntactic plane. There is a general agreement among the great majority of writers on information languages that they are genuine language
systems and that similar descriptions in terms of vocabulary, syntax and semantics are appropriate (Foskett, A. C., 1982, p. 114; Soergel, 1974, p. 28).

2.2.1 Information Language vs. Natural Language

But before we proceed further, we must bring out the important differences between ILs and NLs. These are (cf. Hutchins, 1975, pp. 7-9):

(a) at the formal level, while in NLs the written forms are secondary to the vocal forms and are usually derived from them, in ILs it is the written forms that are basic [1];

(b) at the semantic level, ILs differ from NLs in standardizing the vocabulary through control of homonymity and synonymity;

(c) whereas NLs can function as their own metalanguage, ILs cannot;

(d) ILs are artificial languages -- the governing rules of which are de jure as opposed to those of NL which are de facto;

(e) artificial languages such as ILs, are designed to fulfill a particular function, whereas NLs do not have specific functions, they function in many different contexts and for many different purposes.

Some of the above statements need further qualification, especially (c) and (d). A 'metalanguage' is defined as the language used to talk about another language, the object-language. The English language, for example, can be used as a metalanguage to talk about chemical symbols, the object-language of chemistry, e.g., the sign 'NaCl' means sodium chloride. But English, like other NLs, can also be its own metalanguage, e.g., the word 'sodium chloride' means 'common salt'. But like other artificial languages, ILs cannot function as their own metalanguage. Another language must be used, either a natural language or, more restrictively, another artificial language, e.g. symbolic logic. It was Gardin (1969), who introduced a differentiation which is now generally adopted in France: he said that ILs
are part of the larger class of semantic 'metalanguages' oriented towards application in the field of retrieval of scientific and technical information. Following Gardin, Nöel (1972) proposed a metatheory of linguistics and information science, with a metalanguage having all the properties of a classification schema. The term 'metalanguage' specifies a 'public' metalanguage, such as a document classification system, as distinguished from the 'object' language represented by the documents. Montgomery (1972) agreed with Nöel's definition of a metatheory relating linguistics and information science. But her concept of a metalanguage involves use of a document classification schema as a basis for elaborating a metalanguage specifying a subset of the encyclopaedia, rather than as a metalanguage in itself. Obviously, one can argue that ILs are metalanguages insofar as they are used to talk about (NL) texts of documents. But even if this is true, it is still not the case that ILs can be their own metalanguage, i.e., 'a metalanguage of a metalanguage' (Hutchins, 1975, p. 7).

Information languages are clearly different from natural languages, and seem to fall in the category of artificial languages. In contrast to NL, such artificial languages have been designed for a specific purpose or to fulfil a number of functions. The rules governing the use of NLs such as English, are de facto, i.e., the way most people produce or interpret English utterances and the way it is known by the ideal speaker-hearer, rather than as it 'ought' to be. Whereas, the use of artificial languages such as symbolic logic, is completely determined de jure. Users' must follow the rules prescribed for their proper formulation and interpretation; deviation results in loss of comprehension. For NLs, the breaking of de facto rules does not result into any such drastic consequences; in fact, the rules are frequently broken with little loss of meaning (as in poetry). Or, seen from another angle, NLs and ILs should be distinguished on the basis of
their 'acceptability' criteria (Svenonius, 1979, pp. 63-64). For a NL, an acceptable sentence is one that might be produced by a native speaker of the language, which is a good criterion because it is extralinguistic and independent. But it is difficult to apply such notions of acceptability in an IL, because it is designed by a small group of people for a very large public. It is this small group of people who are seen as competent to judge whether an index term is acceptable or not. Information scientists have also failed to establish any independent, extralinguistic criteria. NLs do not have specific functions, they function in many different contexts and for many different purposes. This functional aspect makes them less efficient as opposed to the artificial languages, which are designed to be more efficient than the former.

The characteristic structures of a language are determined largely by the function(s) it is designed to perform. Among many functions of NL, it is only the descriptive function, viz., the communication of information, which is usually attributed to artificial languages, and very often they are limited within this function to specific facets. For example, ILs concentrate on a particular channel of communication, i.e., documents. Whatever their particular function, artificial languages are designed to be more efficient than NL, in the manner they reduce, or even eliminate, the redundancy and ambiguity prevalent in the latter. Artificial languages seek to 'normalize' NL semantics by reducing the multi-expression concept (characteristic of NL) to a single expression concept. Salton (1968) calls it 'language normalization': "If the natural language is used as a primary input to an information system, any content analysis system must include methods for consistent language normalization". The most obvious way to achieve such normalization is to create either special symbols for the purpose (as in chemistry) or to develop a restricted form of NL, free of synonyms, homonyms, ambiguity and redundancy. ILs use both these methods.
Classificatory languages such as the Dewey Decimal Classification, Library of Congress Classification, etc., use special notations, whereas indexing languages such as, pre- and post-coordinate indexes, subject heading lists, thesauri, etc., use a 'standardized' version of NL. This process of normalization or standardization is a common feature of all artificial languages. ILs differ from other artificial languages in the particular functions they have to perform and in the influences that certain pragmatic factors have upon their structures as language-systems. For instance, the primary function of ILs is to act as channels of communication between documents and potential readers, but unlike many other artificial languages (such as chemistry, music, etc.), nobody expects the readers to learn the IL. Pragmatic considerations such as, an IL should impose an organization of documents and their contents that is generally acceptable to the great majority of the users for most of the time, play a large part in the design of any IL.

2.3 Universals in Information Languages [2]

The history of classification theory is to a large extent the history of a search for the general principles of organization and arrangement that can form the foundations of universally valid and acceptable systems of classification. The search is a natural outgrowth of the basic purpose of ILs, the normalization and standardization of communication: the more nearly the fundamental structural principles of ILs conform to the mode of thinking, writing and understanding of the majority of their users the better they can perform their functions.

With these aims in mind classificationists have searched for universals in many fields of study: from logic and biological taxonomy are derived most of the basic principles of classification (such as, the 'whole-part' relation, the 'genus-species' relation), from the philosophy of science many
ILs have taken the skeleton of the main classes (e.g., Dewey Decimal Classification from the Baconian scheme), from general systems theory the Classification Research Group (UK) has adopted the 'theory of integrative levels' (Foskett, 1961; Austin, 1969), in developmental psychology Farradane (1966) has found the basis for his relational indexing. In searching the 'hidden roots of classification', Ranganathan (1967a, p. 408) suggested that there may be an "Absolute Syntax" for subject descriptions: a sequence in which ideas "arrange themselves in the minds of a majority of persons" regardless of which ordinary human language they speak.

Similarly, during the last decade or so, there have been attempts to find a unified theory of indexing. Because, a complete and commonly accepted theory of indexing, once created, would be a central theoretical construct in information science (Travis and Fidel, 1982). A synthesis of compatible theories may prove to be the first step in establishing a comprehensive indexing theory. In 1977, Borko described indexing theories formulated by Jonker, Heilprin, Landry and Salton, and concluded that although some of them needed to be validated, the "remarkable degree of congruence among these theories ...is indicative that a comprehensive theory of indexing may not be far in the future" (Borko, 1977, p. 365). Unfortunately, no such attempt has been made so far. Rather, several theories have been put forward by researchers from time to time: theory of probabilistic indexing by Maron and Kuhns, Swets, Robertson, etc.; utility-theoretic indexing by Cooper (for general discussions on both these theories and their later developments, see Maron, 1978); axiomatic theory of indexing by Fugmann (1979, 1980); and so forth. However, we are not so much interested in the indexing process and the characteristic of indexing vocabulary, as we are in the nature of the languages used to perform the task of indexing.
2.3.1 Linguistic Universals in Information Languages

Although, many ILs use NL words for their purpose (such as, majority of indexing languages), information scientists have rarely looked for universals in natural language. But this is true only of surface forms in NL. Many linguists believe that at a 'deep' level there are universal features of NL. If this is so, then there are obvious implications for the construction of ILs having some claim to universal validity.

In recent years, there have been several attempts to prove or disprove the theory that search for universals in ILs is essentially parallel to search for universals in NLs. The issue is a multi-faceted one. First, there are those 'non-believers', who assert that an IL differs from NL to such an extent that one cannot expect linguistics to provide adequate explanations of the ways in which IL functions. The group, itself, is divided into two sub-groups, according to the reasoning they provide. One group believes that "Language (NL) is not a medium of expression developed specially for exact and consistent transmission of meaning (as required in ILs)" and these "inconsistencies of language present a formidable barrier" to the forming of universally acceptable structure of ILs (Farradane, 1967, p. 298). NLs are so varied that, "the number of Linguistic Syntaxes for the name of a subject, in the different natural languages all taken together, can become as great as factorial n, where n is the number of kernel terms in the name of the subject" (Ranganathan, 1967a, p. 409). The workers in this group support the use of traditional library classification schemes using symbols, to represent concepts and their interrelationships rather than ambiguous NL expressions. In contrast, the other group regards linguistic explanations as over-refined for the purpose of IL description. They are especially concerned with automated language processing and/or automatic indexing. Workers involved in these researches (e.g., Sparck Jones, Kay, Salton, Montgomery, etc.), feel that the link between linguistics and information
science remained unexploited, because "an explicit knowledge of how human beings receive and transmit information was practically unnecessary so long as information processing operations such as indexing were performed by humans" (Montgomery, 1972, p. 195).

Second, there are those few 'faithfuls', for whom it is a plausible assumption that, should universals of language exist, they may be more easily detectable in an IL than in NL, because, ILs in general are highly structured and condensed (Bivins, 1978).

In between comes the third group, whom we can call the 'moderates'. According to this group, the ways in which subject statements (e.g., titles, subject headings and index entries) convey their meanings cannot always be entirely explained through paradigms developed for the analysis of NL. Two main arguments were put forward to support this view (Austin, 1982b, pp. 130-131):

(a) the linguist is concerned with a wider spectrum of expressions than those found in subject headings, indexes, etc. The latters' are actually formalized subsets of NL, or paralanguages. Consequently, some but not all linguistic theories have relevance for the documentalist;
(b) more strongly, subject statements should not be regarded as a kind or subset of NL utterances, but need to be seen as expressions of different kind, constructed in accordance with their own rules, and therefore requiring their own explanations [3].

The works of Lynch, Coates, Austin, etc., can be cited as examples of indexing languages subscribing to this line of thinking.

Among ILs, the classificatory languages seem to be least dependent on NL principles, if only because concepts are expressed by symbols rather than words. Class- and hierarchy-building and establishment of categorical relationships among concepts were the preoccupation of classificationists.
rather than organizing the concepts deliberately into NL syntagms. But, it is possible to trace certain elements of NL construction (e.g., in its passive declarative sentence form) in classification schemes too (as in Vickery's (1975) "Standard order" -- Thing (Product) - Part - Constituent - Property - Patient - Action - Agent), especially when schedules deal with concepts relatable as object, action, etc. To many classificationists this may appear to be nothing but nostalgia. On the other hand, indexing languages such as, post-coordinate indexes, have given priority to semantic relationships, implied in the terms acting as retrieval keys. Boolean functions (e.g., AND, OR, NOT) are used to link terms at the search stage, which are best regarded as logical rather than syntactical relationships. Despite the fact that, such systems performed fairly well in natural and physical sciences (where the relations are fairly straightforward), problems arose when attempts were made to use them in the social sciences, where a set of concepts can be interrelated in several meaningful ways. Solutions in the form of role-indicating words (e.g., EJC (Engineers Joint Council) and WRU (Western Reserve University) 'Roles'), attachable to indexing terms, were introduced. Some saw these roles as parallels to case marking systems in NLs (e.g., Costello, 1964). But, it should be kept in mind that, for both classification schemes and post-coordinate indexes, these are examples of post hoc explanation, not the description of any ad hoc attempt.

The other category of indexing languages, viz., pre-coordinate indexes, try to resolve the ambiguities faced by post-coordinate indexes, by organizing terms into meaningful subject statements, so that their entries come closer to NL. Therefore, it might be conjectured that, they are the systems (apart from automated indexing systems) most likely to benefit from the application of linguistic universals, if any. Here again, most systems have possibly evolved in a pragmatic fashion, without an obvious need for explicit reference to general linguistic principles. But, it could be
observed that, intuitively, workers were applying linguistic tools (e.g., word order in passive sentences, prepositions, conjunctions, punctuation marks, etc.) in the formulation of index entries. Coates (1960) proposed the use of prepositions as tests of concept relationships during the stage of subject analysis, especially in systems where the order of nouns in index entries is intended to reflect their relative significance. His table of twenty dyadic relationships, each accompanied by its commonest English preposition, has been used in *British Technology Index* (BTI) to determine the order of terms in index entries. Lynch and others (Armitage and Lynch, 1967; Lynch, 1969), in devising the Articulated Subject Index (ASI) to *Chemical Abstracts*, noted the importance of, firstly, the formal order of noun or noun phrase components, for the quality of entries; and, secondly, that of prepositions, connectives and punctuation marks in marking the boundaries of these components in such a way that the parts can be easily re-arranged into a single title-like phrase. Austin's PRECIS was also developed during the late sixties in an attempt to avoid the fallibility of the then existing schemes of general classification for consistent retrieval of information and their unsuitability for computer manipulation. All these came into light due to the Classification Research Group's (UK) work to review the need for a new general bibliographic scheme during 1964-1968 and the involvement of British National Bibliography (BNB) (absorbed into The British Library in 1974) with the UK-MARC project (Foskett, A. C., 1982, p. 254). There was no doubt in Austin's mind about PRECIS' legacy to classificatory principles, especially to faceted classification, when he says "PRECIS is a direct descendant of faceted classification, and "the design of an indexing system which is based upon organized word strings must take some account of the principles which have been developed to regulate the citation order in faceted classifications" (Austin, 1974b, 53). But, certainly, "This is not to say that these principles can be applied with
equal effect to strings of terms selected from natural language". Later, partly due to the success with which PRECIS was applied in non-English languages and partly to avoid suggestions from some quarters (Langridge, 1976) to establish its resemblance with faceted classification schemes, led Austin to seek a new basis for explanation of the general principles working behind PRECIS. As Austin (1982b, p. 1) himself said, "a third important factor which has frequently been overlooked by writers on indexing and classification, ... was an increasing tendency, from the earliest trials with PRECIS, to abandon what might be called a classificatory approach to the organisation of terms in index entries, and adopt instead an order which calls for an explanation in terms of grammatical categories and relations" (italics mine).

2.3.1.1 Evidence from Theoretical Linguistics

Language (or linguistic) universals can be defined as aspects of language which are common to all languages, no matter whether these be universals of structure or of the meaning content of a language. The question, whether there are universals in natural language or not, has generated considerable controversy among linguists in the past. Some linguists even deny the existence of the so-called linguistic universals and claim that human languages have no distinguishing characteristics as such (Kay and Sparck Jones, 1971, p. 143).

Most of the attempts to date to discuss universals have been syntactically-based linguistic studies. In the past two or so decades, there has been a discernible trend towards recognising that such universals as may exist would necessarily be semantically-based. That is, it is the meaning, or even more abstractly, the conceptual processes which are being discussed (Bivins, 1980a, p. 55). A notable exception to this generalization are the many Soviet linguists, who have a unique tradition in lexicology and related
semantic studies equally as important as phonology and grammar. Here again, there is an ongoing debate over the role of syntax versus that of semantics in the explication of universals, and the two sides have been conveniently labeled as 'Transformationalists' and 'Lexicalists', respectively (Montgomery, 1972, p. 210). The issue remains a much-debated and un-resolved one till this date (Warner, 1987, p. 85).

The first step towards the recognition of general linguistic principles is usually traced to Ferdinand de Saussure, who introduced and anticipated a number of important distinctions in linguistics. One such distinction he introduced (as noted in section 2.1), is between syntagmatic and paradigmatic relationships. He also anticipated distinctions between the notions of 'deep structure' and 'surface structure' of language, proposed by later linguists. The development of an 'ideal language' in which each simple idea is represented by a single symbol and an algorithm provided for their combination to represent complex ideas was, a cherished dream of Leibniz. In his classic work on language, Whorf stated that, every language contains terms that have come to attain cosmic scope of reference that crystallize in themselves the basic postulations of an unformulated philosophy, the examples of which are words, such as reality, substance, matter, space, time, past, present, future, etc.

However, the true beginning of the search for linguistic universals began with the school of transformational-generative grammar, led by Chomsky (1957, 1965). Chomsky postulated that what is common in a language structure is the underlying meaning of a sentence. That is, all sentences have a fundamental 'deep' structure upon which transformations can be performed in order to create variant surface forms of sentences. To explain this phenomenon, Chomsky quoted an insightful remark by Humboldt, who suggested, in 1836, that language "... makes infinite use of finite means". For
example, the sentence "The librarian gave the book to the reader" can be transformed by the appropriate rules into the passive form, the question form and so on. Chomsky further speculated that humans are born with the ability to perform these transformations, and that there is a genetic programming for this aspect of language. As supporting evidence for this view, he cited the human ability to learn so rapidly to use language, and, what is more important, to comprehend language. In addition, humans, from the earliest days of speech, are endlessly creative in language terms. We can, at will, both produce and comprehend sentences never before spoken or heard. Thus, the universal here is that of the innate knowledge of the deep structure of language and the ability to perform transformations on it. Names like, Katz, Fodor and Postal, easily come to mind as members belonging to this school.

McNeil commenting on Bailey's work says: "Since innate ideas are not arbitrary, deep structures are universal among languages. In Bailey's theory child and adult speech converge beautifully at the most crucial level -- at the level of the deep structure where meaning is organised -- and diverge elsewhere, at the level of sound". Birnbaum suggested a multi-layered syntactic structure between the deepest of the deep structures and the surface structure.

In Chomsky's transformational-generative grammar, categories in sentences are named, e.g., verbs, nouns, verb phrases, noun phrases, and the like. A number of linguists held that Chomsky's theory did not go far enough in providing for the semantic element in language. The introduction of 'deep cases' as linguistic universals is usually attributed to Fillmore (1968), who accepted transformational-generative grammar as his starting point, but pointed out that transformations cannot be explained adequately without reference to inter-concept relationships more specific than those between noun phrases and verb phrases. According to Fillmore, it is the relations among words (phrases) in sentences which are of prime importance. If
sentences are analyzed for deep structure, then the case relationships appearing in the deep structure are considered central. An example may clarify this: "The teacher distributed the course-work among the students". In this sentence, the deep structure involves the four words, viz., teacher, distribute, course-work, students. The central action is the verb 'distribute' around which the rest of the sentence revolves (Figure 2.2). The object of the action distributing is 'course-work'; the agent is the 'teacher', while the benefactor (recipient) of the action is 'students'.

![Figure 2.2 Case Relationships](image)

Fillmore introduced the term 'deep case' as a generic to cover a range of syntactical situations which may or may not be manifested as surface cases. He offered a basic set of deep cases as linguistic universals, the more important of which, from an indexing point of view, are agentive, instrument, dative, factitive, locative, objective, benefactive. This theory is fundamental to a linguistic explanation of Austin's PRECIS.

### 2.3.1.2 Cognitive Aspects of Language Processing

Both of the previous theories (i.e., Chomsky's transformational-generative grammar and Fillmore's case grammar) have concentrated more on the syntactic than the semantic level, despite Fillmore's intent to extend the 'meaning' aspect of transformational-generative grammar. A whole range of inquiries from areas such as, artificial intelligence, machine translation, cognitive psychology and linguistics, have given rise to the prevalent notion that the cognitive aspects of language processing need to be taken into account in
attempting to develop any further theories of language. A central question in psychology (and to some extent linguistics) and one which has been argued extensively in the history of philosophy, is this: is meaning pure thought, and is the rest the formal language we speak? Several attempts have been made to answer this question during the second half of this century.

Ideas are largely products of intellectual activity, which is known to be controlled by brain. There is considerable similarity in the structure, and therefore, in the functioning of the brain in a majority of normal human beings. Thus, a majority of normal human beings have more or less a similar mode of thinking and learning -- that is, in forming ideas and in combining them to build knowledge-structures. It is further stated that biologically man has not changed to any appreciable extent since the emergence of Homo sapiens; for, the structure of the genetic material has not appreciably changed since then -- that is, for some 500,000 years -- although we have changed culturally. Therefore, the probability of a sudden change -- that is, a mutation -- in the mode of thinking and learning of a majority of normal human beings in the immediate future is quite low (Neelameghan, 1971, p. 325). Hence, if the cognitive processes are biologically similar and expression of thought into some kind of deep structure of language can be modelled, then the resulting universals might be better suited for application into information languages. What follows is an outline description of some such theories.

A. Approaches to Thought and Language

Object-referent theory. This is one of the oldest theories of meaning. It's invalidity is rather easy to demonstrate. It is evident that there are many names for things which are easy to see: libraries, students, book-mobiles, and so on. In attempting to generalize about natural languages, a referential theory of meaning inevitably runs into difficulty for the reason
that many words lack real-world referents. For instance, it is difficult to imagine what is named by abstract words such as, attitude, skill, efficiency, etc., since these correspond to no physical entities in the real world. According to Svenonius (1979, p. 74), "To meet criticisms like this those who endeavour to maintain a consistent referential theory of meaning are obliged to invent perceptual or conceptual constructs to serve as referent for abstract words -- in effect, a Platonic heaven".

**Behaviourist approach.** Psychologists belonging to this group believed that, thought is the movement of vocal musculature. If one cannot speak, *ipso facto* one cannot think. However, tests carried out on volunteers, immobilized by a certain drug injected in their body, reported that they were able to think and mentally complete several pre-assigned problem-solving tasks. This theory, in terms of language, and by extension, of meaning, would hold that the meanings of words are acquired through imitation, practice, and reinforcement. Children do imitate some of the sounds they hear, but, at the same time, they also produce many of their own which has no parallel in adult speech. But the behavioural theory has failed to explain the reason behind the child's ability to produce and understand sentences which he/she has never heard before.

**The cognitive development viewpoint.** In this view, held by Piaget, Bruner, Vygotsky, among others, language and thought influence and reflect one another. Thought, however, is the pacemaker, which is integrated through the developmental process. Piaget and Inhelder pointed out that "It is possible to show the similarity between Piaget's description of sensory-motor structure and Chomsky's deep structure of language". But, the major difference is that, Chomsky believes that humans are born with a genetic programming for certain language universals, and, thought and language are entirely separate processes. Piagetians believe that language is grafted on
to thought and there is nothing such as innateness of language. Many
linguists and psycholinguists concerned with the interconnections between
language and thought, hold an amalgamation of both the views.
Psycholinguists take an experimental approach, as opposed to Chomskyans who
are fundamentally theorists, as we have mentioned in the previous section.

B. Experimental Theories of Meaning

Semantic features and related hypotheses. Among the experimental theories of
psycholinguists, the most widely known hypotheses are those relating to
semantic features. It requires division of a word into sub-units. The
meaning of a word is not considered an indivisible unit; for instance, the
word 'bat' is ambiguous. It can mean either that which is used to hit the
ball in cricket, or a small, flying, furry mammal. Once qualified with
cricket -- cricket bat -- the ambiguity is eliminated. But, if one says, "I
saw the bat", the meaning is ambiguous. According to Clark, the child
gathers information through his perceptual system to which he relates
possible meanings of words. Children seem to acquire simpler concepts (those
with less features) first, and, conversely, tend to assume simpler meaning
when using the more complex term. A number of related studies evolved after
the semantic features hypothesis. Their findings may be summarized as:
(a) terms develop in children systematically as a set, rather than as
individual words, i.e., the words do not exist in isolation;
(b) particular concepts are central to the development of terms;
(c) focal colours are defined by unknown, as yet, perceptual and cognitive
factors common to all human beings; and
(d) the actual or proposed development of hierarchies is dependent on the
features of words however they may have been obtained.

Sentence processing (information processing) approach. Theories invol-
vving semantic features, while perhaps considering an overall cognitive
development, concentrate on the whole on individual lexical items. But, a radically different approach was followed by Bransford and others. They conducted a number of experiments which involve processing of information from sets of sentences, rather than words. Results so far seem to indicate that information is stored in human memory in a non-linguistic fashion as the 'gist' of the sentence(s). The results hold for both adults and children. One such experiment showed that, subjects are most confident that they heard a sentence which combined all the ideas of the initially presented several short sentences. This produces the evidence that some sort of abstraction of the gist of the sentence takes place in the human brain.

The two theories mentioned above do not necessarily contradict one another. It may be that stored in the 'memory' are features and relationships between words; it may be that actual sentences are stored in an abstracted fashion in some other area, and in some other way. The state of our knowledge in this area is still in flux.

C. Artificial Intelligence Approach

Investigators in artificial intelligence have used a different approach, that of simulation or model-building, to represent the process of concept formation in human brain. They assumed that, it is more than structure, perhaps more than meaning, that needs to be tapped in order to devise an automated method for dealing with language. Following are some examples of the work carried out in the area of modelling of conceptual processes:

Quillan's semantic memory model. Quillan's central concern is how semantic information is stored within a person's memory. He proposed a complex model of what can be termed a 'spreading-activation' network. Words do not exist in isolation, but are part of an interrelated network, parts of which are variable. Quillan initially constructed a small system consisting
of a network of some 60 terms. These terms were empirically defined and quasi-hierarchically structured. A series of experiments confirmed that there are levels of relationship within the network, which prove that, information which is stored directly with a term, is recalled at a faster speed, as opposed to that information which is stored in a remote location. A very typical example quoted to explain this point is that one can respond faster to a question such as "Is a canary yellow?" than to a question such as "Is a canary mammal?".

**Episodic memory model.** The model proposed by Norman and Rumelhart is more comprehensive than Quillan's and relates the existence of 'episodic' memory stores. An episode is a series of events or actions. Episodic memory lacks universality, that is, all persons are likely to have unique experiences which are stored in a holistic manner in the memory.

**Frame theory.** The various studies in linguistics, cognitive psychology and artificial intelligence outlined so far, have one way or another attempted to derive some form of universality in either human thought processes or languages, or both. But in Bivins' (1980, pp. 58-59) opinion, "the meaning of word is more than its structural relationship, more than its case relationship, and more than a collection of components or features. It consists of word 'meanings' certainly, in a dictionary sense; it also consists of facts, information, experiences, in an encyclopedic sense; furthermore, it also consists of what might be called 'action sets' which are performed by persons as a ritual of a sort (say, greetings)... Finally (although the list is probably not complete), one might need, as part of a total meaning, such as problem-solving approaches and other types of heuristics. What these notions involve, as a whole, is a type of conceptual frame".

The frame theory can be seen as a further extension of case grammar.
but owes its origins to artificial intelligence. In contrast to the word
and/or sentence-based theories of Chomsky and Fillmore, the frame theory is
contextually based. Words do not have meaning in isolation, but depend, for
their meaning, on their experiential context. It is difficult to single out
the universal aspect of a frame theory, but one might postulate the
existence of a prototypical, 'universal' frame which would be common to all
humans, e.g., the best instance of a 'dog'. Though, Bivins (1979) seems to
be aware of the obvious flaws of this method, she thinks that something akin
to frame theory is needed in furthering information retrieval development.

But the artificial intelligence approach has its share of criticism
too. In Sparck Jones' (1979) opinion, artificial intelligence meaning
representations are different in kind and not merely in degree from document
retrieval descriptions and that in current information retrieval it is
correct to think in terms of 'aboutness' distinct from meaning
representation. She believes that, more sophisticated information retrieval
systems will depend on linguistic techniques of meaning representation.

2.3.2 Evidence from Related Disciplines

A. Biocybernetic view

In his book on systems philosophy, Ervin Laszlo mentioned the 'basic modes
of thinking' and realized that, all men, regardless of the culture they
happen to belong to, have basically similar nervous systems, are equipped
with analogous sense receptors, command like patterns of response, and use
patterns of thought which obey very similar laws or regularities. In other
words, there appear to be some 'universal' traits underlying cultural
cognitive relativities: Chomsky could locate 'linguistic universals' and
Kluckholn discovered a number of 'universal categories of culture'.

B. Evidence from Symbolic Logic

Suzzane Langer pointed out that the psychological context of our thoughts may be private and personal. Therefore, two persons talking together might have different perceptions of the same thing. They are then said to have different 'conceptions'. But, if they let it be understood to each other, then their respective conceptions embody the same 'concept'. A concept is an abstracted form. Abstraction is the consideration of logical form (structure) apart from content. In this sense, the sentence processing approach could be described as an extension of this approach.

C. Logic of Exposition and Linguistic Syntax

According to Arturo Rosenbleuth, there is a basic difference between mental events and the correlated neuro-physiological processes. For example, someone has presented verbally a specific topic on three different occasions in three different languages, viz., English, German and French, respectively. Although, the neuro-physiological correlates corresponding to the logic of his exposition might be similar or identical in the three cases, clearly those corresponding to the selection of words and their syntactical organization would be absolutely dissimilar. Thus, this theory supports the findings of the Piagetians in particular, that, thought precedes language, and the cognitive psychologists in general, that, in deriving universals thought processes are more dependable than linguistic expressions.

D. Syntax of Knowledge and Epistemics

Meredith suggested the existence of a 'syntax of knowledge', which, even if not entirely independent of the particular languages, can and does, in practice, follow its own course alongside the syntactic sequence of
language. As a proof, he mentioned that, at a multilingual conference, with a host of disciplines, experience and thought, the translators hardly find it difficult in transforming, almost instantaneously, the most elaborate syntactic forms of one language into the quite different forms of another whilst preserving the essential structure of information and conceptualization in the speech. Even though the syntax of language cannot be entirely separated from the syntax of knowledge, we can pragmatically separate them by treating the one as a temporal sequence (by the sequence of words in the sentence) and the other as a spatial structure (a geometric knowledge of evidence spread out in space or held in memory).

2.3.3 Universal Forms and Subject Representation

Parallel to the search for universal linguistic forms such as those expounded by Chomsky, Fillmore, and others, and recognition of similar universal traits in human memory and process of knowledge acquisition, steps towards the formulation of a generic framework for structuring the representation of the name of a subject for the development of classification schemes and indexing languages were also investigated. Such universals are being arrived at and used in various other areas dealing with information storage and retrieval. For example, in the area of data modelling now the basic problem is to identify the world as a domain of objects with properties and relations (Biller and Neuhold, 1978). Although, classificationists have been involved in finding such a generic framework since the days of Dewey and Cutter (as mentioned in the beginning of section 2.3), it was not until the 1950s that such need was explicitly expressed by the Classification Research Group (1955), in the United Kingdom. By that time, it was realised that existing enumerative schemes of classification were falling apart in their attempt to keep pace with the 'information explosion', and especially, these were inadequate in specifying complex
subjects. As an alternative, a new approach was advocated: a turning aside from enumerated lists of subjects, and the adoption of freely faceted principles and analytico-synthetic approach. This resulted into an upsurge of classification schemes, mostly in special subject fields, based on what is now epitomized as the technique of 'facet analysis'.

2.3.3.1 Ranganathan and the Indian School

However, the above realization was not entirely new to the library profession. As early as the 1940s, the use of categorization of component ideas forming the name of a subject into a few Fundamental Categories, and defining in order of these categories to form a 'logical classificatory language' resulting in 'faceted' library classification schemes were developed in India by Ranganathan (1933, 1937). But, Ranganathan and his work remained almost unheard in the West (Foskett, D. J, 1982), till B. I. Palmer went to India during World War II and brought it to the knowledge of the profession at large. But, first we will have a look into the developments that took place in India, especially at the Documentation Research and Training Centre (DRTC), Bangalore, under the guidance of Ranganathan.

It was realized that, the order of component ideas denoting the different categories in the name of a subject as prescribed in the PMEST formulae is a context-dependent order. More specifically, it is a context-specifying order (Devadason, 1983, p. 23). Every component category sets the context for the next and following ones. Also, in this classificatory language, every category should explicitly have the corresponding superordinate component ideas preceding it. The reason for fixing the superordinates before the component elements concerned is to make the component elements denote precisely the ideas they represent. Further, it
has been conjectured that the order (syntax) of representation of the component elements in the name of a subject as prescribed by the principles for sequence -- facet sequence -- is more or less parallel to the 'Absolute Syntax' -- the sequence in which the component ideas of subjects falling within a subject field arrange themselves in the minds of majority of intellectuals (Ranganathan, 1967a; Neelameghan, 1971). This idea of absolute syntax is similar to Chomsky's (1975, p. 4) idea that, "more intriguing is the possibility that by studying language we may discover abstract principles that are universal by biological necessity and not mere historical accident, that derive from mental characteristics of the species". Not only linguistic universals, but also, similar universals were evidenced in other fields of knowledge, e.g., biocybernetics, philosophy, epistemology, etc. (as described in section 2.3.2). If the syntax of the representation of component ideas in the name of subjects is made to conform to, or parallel to, the absolute syntax, then the pattern of linking of the component ideas -- that is, the resulting knowledge-structure -- is likely to be:

(1) more helpful in organizing subjects in a logical sequence for efficient storage and retrieval;

(2) free from aberrations due to variations in linguistic syntax from the use of the verbal plane in naming subjects; and

(3) helpful in probing deeper into the pattern of human thinking and modes of combination of ideas (Neelameghan, 1979, p. 170).

It is interesting to note that it has been realized that Ranganathan's fundamental categories of Personality, Matter and Energy are "general categories building the system's structure as a spatio-temporal neighbourhood relationship" useful in deriving meta information for a process of automatic semantic analysis too (Ciganik, 1975). Recent research on the psychological and linguistic aspects of PMEST structure (Iyer, 1984) and the ability of such a structure to communicate information with least
distortion (Raghavan, 1985), suggests that the PMEST structure has psychological roots, in the way human mind thinks and organizes concepts. Therefore, this can constitute a cognitive model which can be effectively used in information retrieval (IR). Iyer (1986) has also shown that, Ranganathan's PMEST model can be effectively used in developing the network representation of the user's problem statement, in evolving standardized search strategies to resolve anomaly (as viewed in Belkin's (1980) ASK (Anomalous State of Knowledge) hypothesis), thereby minimizing the intervention of the search intermediary.

Consequent to the development of techniques for structuring of subjects and for classification of subjects, several experiments were conducted at the DRTC to use them for subject indexing, thesaurus construction, formulating search queries for computerized databases, etc. (Devadason, 1986b). Till about 1935, the processes of classifying the subject of a document and of preparing the subject headings for it were considered as if they were independent and mutually exclusive. But with the development of chain indexing (originally called 'chain procedure') by Ranganathan (1934, 1938), it became obvious that subject headings can be constructed by a translation into a meaningful representation of the class numbers in the verbal plane. Large scale use of chain indexing in big university library catalogues as well as commercial publications, such as British National Bibliography and British Technology Index, bear testimony to its helpfulness in structuring a subject in the verbal plane and in formulating subject headings in a consistent, systematic, and economical way (Neelameghan, 1975). But, as early as the beginning of the sixties, it was becoming apparent that chain indexing was not always helpful in subject specification and in its service to the user (Sweeney, 1970). This culminated in BNB's rejection of chain indexing as the tool for its alphabetical subject index generation and adoption of PRECIS as a favoured
replacement. Also, the latter claimed to be better suited to a mechanized environment. However, it appears that one particular work by Ranganathan went almost unnoticed in the Western hemisphere, especially in the United Kingdom, which showed that, "the choice of the name of subject of a document and the rendering of the name in the heading of a specific subject entry can be got by facet analysis based on postulates and principles" and also pointed out that "using facet analysis for subject heading does not amount to using class number" (Ranganathan, 1964, p. 109). Further research into the fundamentals of subject indexing languages resulted in the development of a general theory of subject indexing language (Bhattacharyya, 1979a, 1980). The POstulate-based Permuted Subject Indexing (POPSI) language was developed through logical interpretation of the 'deep structure' of subject indexing language forming part of the general theory of subject indexing language (Bhattacharyya, 1979c, 1981). According to Ranganathan, there are only five basic facets (categories): Personality, Matter, Energy, Space and Time, to which one can add the Basic facet or the Discipline facet [4]. These have been abstracted and interpreted recently to be contained in the four facets or Elementary (fundamental) Categories: Discipline, Entity, Property and Action, and a special component called Modifier, which constitute the 'deep structure' of subject indexing language (Bhattacharyya, 1979b, 1979c). It also makes use of the concepts of 'Base' and 'Core' to be used as the first context-specifying category and second context-specifying category, respectively, to enable 'need-oriented' indexing (Bhattacharyya, 1982; Bhattacharyya and Chandran, 1983). The computerized Deep Structure Indexing System (DSIS) is based on the 'deep structure of subject indexing language' (Devadason, 1986a).

2.3.3.2 Classification Research Group (UK)

We have already mentioned CRG's acceptance of freely faceted principles and
analytico-synthetic techniques as the basis of all methods of information retrieval. In the next step, CRG proposed that the concept of main classes should be abandoned as the basis for a library classification, at least during the initial stages of its design. It directed its attention to the 'universe of concepts', i.e., those discrete ideas which constitute the components of the name of a subject in all fields of knowledge, rather than to the 'universe of subjects' and its organization into disciplines (Austin, 1976). This change of direction had been noticed by A. C. Foskett (1982, p. 234), who pointed out that: "Starting with the theories of analytico-synthetic classification developed by Ranganathan, the Group has moved forward in a rather different direction from Ranganathan himself". That means, each concept would be assigned to a category according to its meaning, and without being associated with a discipline from the outset. It would then be notated on a once-for-all basis. Compound subjects would be built by synthesis out of these elements, using generalized rules to determine the order in which concepts should be cited. Two different but interrelated lines of enquiry emerged out of this necessity: (a) establishment of a general system of categories; and (b) search for a universal citation order of concepts.

In its first approach to the universe of concepts, CRG employed the same approach as the one employed by earlier classificationsists in their approach to the universe of knowledge. Depending on the human ability to judge likeness/unlikeness, two general classes of concepts were established: (i) entities or things; and (ii) attributes of entities (Austin, 1982a). Entities were divided into two mutually exclusive groups, naturally occurring things and artificial things. The latter class was further divided into: (i) concrete substances and objects (artefacts), such as 'libraries' and 'film'; and (ii) mental constructs (mentefacts), such as 'theorems'. Attributes were distinguished as either properties or activities, and each
of these was further subdivided, a new principle of division being introduced at each step.

Work on the second component began with accepting Ranganathan's PMEST formula as the basic model. But subsequent research proved that, in many respects it was too imprecise for practical use. This applied particularly to the primary facet, 'Personality'. It was pointed out that, the classifier cannot in practice decide which concept should be assigned to the fundamental category Personality, and therefore cited as the first element in a subject statement, until the subject as a whole has already been assigned in his mind to some appropriate class. This means, there is an element of intuitive perception in the recognition of Personality, rather than clear cut rules (Roberts, 1969). Finally, CRG devised a new schema which depicts a mixture of features belonging to two other systems, those of Vickery and Farradane. The order of concepts in a compound subject was generally that proposed by Vickery (1975):

- Things; Kinds; Parts; Materials; Properties; Processes; Operations; Agents.

This formula ensured that concepts should be set down during number building in their order of relative significance, but not necessarily in a 'meaningful' order, nor on the basis of inter-concept relationships. It was Farradane who adopted relational analysis of concept-pairs as the basis of arrangement of concepts in a name of subject. Instead of imposing a predetermined order upon the components of a compound subject, he proposed that, concepts, represented by English terms, are to be connected by symbols called 'relational operators', each of which expressed one of nine dyadic relationships (Farradane et al, 1966). The operator /-, for example, expresses 'Reaction', a group of relations which expresses 'action of a thing or process on another thing or process', and this would be written as a connecting device in the index entry:

Serials /- Acquisition,
i.e., 'Acquisition as the action performed on serials'. In the resulting 'analet', both concepts and their interrelationships were thus made explicit. The user who has mastered the relational codes can then read the analet as a meaningful statement.

But, finally, following Vickery, it was decided that the order of concepts in a compound subject should be determined by the use of a general precedence formula. The syntactic role of every concept except the first would also be expressed by a special symbol written as a prefix to its notation. The roles employed for this purpose were broadly those prescribed by Vickery, together with 'Space' and 'Time' as found in Ranganathan's PMEST formula. The role symbols were written between parentheses (the procedure still followed in PRECIS string writing) and their filing value would determine the order of concepts (PRECIS has abandoned the use of the retroactive order). Another important distinction recognized during the CRG research is that between the substantive and adjectival element in a subject, which was later used as the 'differencing' procedures in PRECIS. For Example, the property concept 'relevance' can be used as it is in a name of subject 'relevance of documents', as well as to denote a class of documents, as in 'relevant documents'. All these developments leading to the generation of a new general classification scheme have contributed one way or another to the final shaping of PRECIS and its claim of universality.

2.3.3.3 Comparative Method of Finding IL Universals

One of the most obvious methods of finding universals is to compare as many systems as possible. For NL this would represent an almost impossible task, since no single effort can claim to have comprehended all existing languages existing in this world. Explicit theorization and testing each individual NL against it might be a viable alternative. Examples of such generalizations
within, say, the context of transformational-generative grammar are the universal status of certain semantic components (e.g., \( \pm \) ANIMATE, \( \pm \) HUMAN, \( \pm \) MALE, \( \pm \) YOUNG, etc.) and of certain 'deep' case relations (e.g., objective, agentive, benefactive, etc.).

As far as ILs are concerned the comparative method is not a practical impossibility, since their number is always within comprehension. But, with the exception of Grolier's (1962) study of 'general categories' and Soergel's (1967) survey of relational categories, very few comprehensive comparisons of ILs have been undertaken so far. An attempt of different nature was, however, made in the Aslib-Cranfield Research Project (Cleverdon, 1967), which viewed IL structure as an 'amalgam of recall and precision devices'. Hypothesized IL Universals were defined in terms of their functional effectiveness, as promoters of recall or precision. But, Hutchins (1975, p. 131) has pointed out that, any hypothesized IL Universals are valid only within the framework of a particular model and do not have much or any validity outside it.

2.3.3.4 Intermediary IL or Switching Language

During the recent past, a new idea has emerged in the search for IL universals, the idea of an intermediary IL or 'switching language' (Coates, 1970; Gilchrist, 1972). It was realized that, compatibility and convertibility of indexing systems and languages are a prerequisite for effective information retrieval on an international scale (Mølgaard-Hansen, 1971). Of course, the stimulus is an economic one: reduction of indexing duplication. Three approaches to compatibility among ILs were identified: (i) establishing a concordance among the ILs used, (ii) the use of a single IL in all systems, and (iii) developing a switching language to move from one IL to another (Wersig, 1979).
At the simplest level where only two or three information centres are involved and only two or three ILs, a bilingual or trilingual lexicon showing how to translate a descriptor in one IL into another IL, is what we need. However, true compatibilities throughout the vocabularies are hardly evidenced. Problems of homonymy, synonymy, descriptor specificity, etc., vary considerably from one vocabulary to the other. Neville (1970), Niehoff (1976) have thrown some light on the methods of reconciling IL vocabularies.

The second alternative approach is to merge the ILs into a single IL for all the participating information centres. But this is possible only when the ILs have the same structural forms, e.g., post-coordinate indexes with controlled list of descriptors. Where the IL structures differ greatly or where there are more than three or more ILs involved the reconciliation approach becomes virtually impossible. Also, the development of a unique common IL may not be feasible because of the variety of uses to which it is put (Vilenskaya, 1977).

Instead of creating as many equivalence tables as there are pairs of ILs between which information is to be transferred, a series of concordances are constructed between each of the ILs and the switching language. The switching language has to operate at a 'crude' level, because, as an intermediary IL it must be as neutral as possible and select the most generic representation of descriptors. This last factor leads to a low general level of indexing specificity on which the intermediary IL would operate. One such switching language suggested by FID and supported by Unesco is Standard Reference Code (later renamed as Subject-field Reference Code) or SRC, which was intended as a "macro-switching code... for linking or switching between relatively broad subject-fields in major information exchange systems" (Lloyd, 1972). One obvious need was the creation and application of a single Broad System of Ordering (popularly known as BSO) which could function as an international switching device. Such a scheme was
finally published (Coates et al., 1978), although its wide use as a practical scheme is still a dream.

It is interesting to note that, despite all these efforts towards establishment of a universal IL, information scientists have so far failed to achieve a truly universal IL. The search is still continuing. Nevertheless, as a result of these efforts our knowledge of IL structure and functions have considerably improved. PRECIS with its language-independent features presented some hope. So does the version of computerized POPSI with its culture-independent syntax of facets.

2.4 A Note on String Indexing

A string indexing language is "a subject indexing language whose expressions, that is, index entries, are multi-term strings which are constructed according to regular and explicit syntactical rules" (Svenonius, 1978). Usually these index entries are computer generated from input consisting of index terms labelled in such a way as to indicate semantic or syntactic categories (facets) as well as relationship with other terms. It is the computer power which made "sophisticated string indexing practicable by taking the burden of generating, sorting, and displaying multiple index entries from the shoulders of human indexers or clerical staff" (Craven, 1985, p. 256). Thus, in brief, a string indexing system is "one in which computer software normally produces two or more overlapping index entries from a single source description of any item" (Craven, 1988, p. 133). In its various forms string indexing has become one of the modern solutions to the problem of handling the mass of information available on the one hand and the fast accessibility of relevant literature on the other (Dahlberg, 1986).

Although traditional pre-coordinate systems began to practice some form of stringing terms in a phrase long ago, it was Ranganathan, who spearheaded
this development. Colon Classification, with its features of facet analysis and sequencing devices, is the forerunner of modern string indexing systems (Svenonius and Schmierer, 1977, p. 338). Craven has divided string indexing systems roughly into three categories according to type of input string they use: (a) phrases in ordinary language, such as various keyword systems, PERMUTERM, ASI, etc., (b) simple lists of terms, such as ABC-spindex, TABLEDEX, SLIC, etc., and (c) strings containing additional codes as computer instructions, such as PRECIS, POPSI, NEPHIS, CIFT, etc. Some other close relatives of string indexing, but certainly outside the mainstream of string indexes (in terms of the following definition), are Kaiser's Systematic Indexing, Chain Procedures, Keen's Universal Index Entry Generator, etc. In recent times, perhaps the best known string indexing system emphasizing coding of input strings is Derek Austin's PRECIS. Similarly, POPSI refers to a highly generalized string indexing procedure developed by Bhattacharyya and others at DRTC, Bangalore, which owes its origin mainly to the General Theory of Classification by Ranganathan and his Indian school.

String indexing systems are mostly the product of two factors: (1) the availability of computers to assist in the production of index displays and (2) dissatisfaction with the inefficiencies, tedium, and inconsistencies of manual methods. In addition to the usual tasks a computer performs in index display production, viz., the sorting of index entries into a recognizable order, the insertion of cross-references from an approved list, and the formatting of the index for display, an additional task is also performed by it in string indexing systems: generation of multiple overlapping index entries from a single input string. For the purposes of this study following definitions of string indexing by Craven (1986, pp. 3-4) have been used as standards: "A string index is a form of index with two main characteristics: (1) each indexed item normally has a number of index entries containing at
least some of the same terms and (2) computer software 'generates the description part of each index entry according to regular and explicit syntactical rules. The description part of a string index entry is called an *index string*; the computer software that produces it, an *index string generator*'.

The rest of the thesis is an attempt to describe and compare two such string indexing systems, viz., PRECIS and DSIS (or computerized POPS1).

Notes

1 Also, "As information handlers, we are primarily interested in the written mode. We deal with the products of the human communication process which are usually expressed in print" (Bivins, 1980b, p. 63).

2 The discussion in this section and ensuing sub-sections owe their origins primarily to the works of Hutchins (1975), Austin (1982b), Bivins (1980b), Neelameghan (1979) and Foskett (1970).

3 This view of indexing languages has been contrasted by Michell (1979), for whom subject-descriptive strings are not systematically different, linguistically, from NLs.

4 The omnipresent Discipline facet in the name of a subject marked the first point of departure in the future work of the Classification Research Group (UK), as described in the next sub-section.
References


Classification Research Group (1955). The need for a faceted classification as the basis of all methods of information retrieval. Library Association Record, 57 (7): 262-268.


CHAPTER 3
AN INTRODUCTION TO PRECIS

It is not as easy as it might seem at the outset, to write a comprehensive introduction to PRECIS. During the past twenty or so odd years of its existence, it has raised considerable interest among information scientists all over the world. The literature abounds (Sørensen, 1979) and is multifocal (Mahapatra and Biswas, 1983). More intriguing is the fact that neither a classificatory explanation nor a linguistic explanation is capable of doing proper justice to a system such as PRECIS, which was born out of classification research, but has sought for universal principles in linguistics. But with a view to future needs (that is of comparing with DSIS, which is essentially based on classificatory principles) a slant towards linguistic explanations may be a better choice [1].

3.1 Background and Objective

The previous chapter (especially section 2.3.1 and 2.3.3.1) has already set PRECIS in the context of its linguistic and classificatory connotations. A repeat seems to be wasteful. However, a few sentences might be enough to set the discussion in its own right. As a recapitulation, it is to be remembered that, PRECIS represents the merging of two separate but related streams of development. The first of these is the CRG's work to review the need for a new general classification scheme and the second is the involvement of BNB with the MARC project. As regards the first, it may be noted that one purpose of the CRG's research and Austin's secondment to the project in 1968-1969, was actually to develop a new classification scheme which might be used for the arrangement of British National Bibliography (BNB) in place of Dewey Decimal Classification (DDC). BNB was looking towards mechanization at the time, and wanted a scheme that would be better suited to this purpose.
than DDC, as well as one that would give a more satisfactory arrangement to the recognition of the two fundamental categories of entities and attributes, the development of a set of relational operators to express relationship between concepts, and a general decision-making model for determining the citation order of concepts in a compound subject. The most important thing which Austin had to bear in mind while formulating the principles for the new system was, of course, the neutrality of the system to various classification schemes. The notion of a neutral word string which can express, in a comprehensive summary form, the subject of a document, and as far as practicable with a natural language order, had an important bearing on the organization of work in the Subject Division of BNB, where the writing of PRECIS strings was first done (Austin, 1974b, p. 78).

The second stream of development which influenced the development of PRECIS was the introduction of MARC (Machine Readable Catalogue) format by the Library of Congress and BNB's involvement in the MARC project. MARC was conceived at the Library of Congress as a procedure for producing catalogue entries directly from machine-readable data. The main advantage of MARC format was in its release from the need to produce the 'main entry' as is done in a conventional catalogue entry. Every element in a full bibliographical description is assigned to an allotted place on the record and identified by its unique field or sub-field code. Thus, the order of elements in the record becomes unimportant, which could be changed at will with the help of simple computer instructions. But soon it was realized that MARC records can also serve as a medium for international exchange of bibliographic information. Since then various national bibliographic centres have implemented similar formats. So there exists a family of MARC formats, of which UNIMARC promises to be the most suitable to achieve this goal.

But BNB's involvement in the production of UK-MARC records was from the
very outset seen as a contribution to an international bibliographic exchange network, and it was only a matter of time before the computer-held data was used in the production of the printed list of BNB. The attempt to produce a standard format created its own problem. The BNB has a classified sequence, complemented by author-title and subject indexes. Before BNB's link with the MARC project, the latter used to be a chain index, geared to the unique version of DDC. Apart from the obvious problems of using this procedure with the then forthcoming 18th edition of DDC, the chain index also proved unsuitable for automation. All these developments led to the formation of a project to devise a new indexing system that would meet the needs of a printed publication produced from computer-held data. This was the research which led to PRECIS.

This research was conducted with the following objectives in mind (Austin, 1976b, p. 5; 1982b, p. 8):
1) The original indexing was to be intellectual, but all subsequent operations, including the generation of all entries and their filing, were to be done by the computer.
2) Each of the entries produced must be co-extensive with the subject at all access points. This is in clear contrast to the chain index (where only the final entry is fully co-extensive) and the subject heading lists (where a given heading may express part of the whole subject).
3) Each entry should be meaningful to the user, preferably without the need for explanation, which in practice means that the language used should be close to natural language ('Public libraries', rather than 'Libraries, Public') and relationships that aren't explicit being made so by use of natural language devices such as prepositions rather than neutral set of symbols.
4) Indexers should, for the sake of consistency and collocation, work within the framework of a common set of indexing rules. These rules should be
applicable irrespective of the subject-field or media.

5) The vocabulary should be natural language based, but allow homonym-synonym control. It should be open ended as well, to admit newly emerging terms as and when needed.

6) The terms in the index entries should be supported by a system of cross-references, generated by the computer from a machine-held thesaurus.

The first version of PRECIS, known as PRECIS I (Austin and Butcher, 1969), was successfully introduced in BNB in 1971. But certain difficulties experienced during its application led to the revision which is generally referred to just as PRECIS and published as the authoritative PRECIS: A manual of... (1st ed.) (Austin, 1974a). PRECIS remained fairly stable since then, until researchers started to experiment with its suitability in languages other than English. As a result, several improvements were drafted (Libri, 1976) and generalized in the second edition of the Manual (Austin, 1984). However, the most detailed explanation of PRECIS' language-independent features is available in Austin's (1982b) thesis. For the purpose of this project the last two sources should be considered as the most authoritative and up-to-date account of the system in all its ramifications. As it has been already opted at the very beginning of this chapter that the discussion will have an overtone of linguistic analysis, no attempt has been made to present the system as such. Besides the Manual (2nd ed.), description of this sort is available in Dykstra (1985), Richmond (1981), etc.

3.2 Nature of and Relationship between Concepts

If the above-named objectives are considered overall, it can be seen that these were mainly concerned with the nature of relationship between concepts. 'Concept' means 'a thing conceived, a general notion'. In PRECIS, as in common usage, the word 'concept' generally conveys a single idea or
unit of thought. For special purposes, however, such definitions lack sufficient precision. For example, a single idea such as 'elasticity' can stand as a unit of thought in ordinary language, but in the context of a specific subject like, economics, the word has no meaning unless it is added to other words like, 'income', 'price', etc., thus making the concepts as 'income elasticity', 'price elasticity', etc., capable of being used as meaningful concepts for indexing purposes. As we will see shortly that, this idea of 'context-dependency' has an important role to play in rendering an index entry intelligible and unambiguous. A 'term' (index term/indexing term) is the verbal representation of a concept and may consist of one or more words, which an indexer ascribes to a document to describe its subject matter. A simple concept such as 'cataloguing' is represented by a term consisting of only one word. But, concepts such as, 'cooperative cataloguing' and 'cataloguing in publication' are represented by compound terms consisting of phrases with more than word. The technique of 'differencing' is based on the recognition of these variant structures of compound terms in natural languages. PRECIS distinguishes between two fundamentally different kinds of relationships between concepts, which can be broadly described as:

1. **syntactical**, i.e., the relationships between the terms in input strings which together express the subject of a document, and in the entries generated by algorithms out of these strings.

2. **semantic**, i.e., those *a priori* relationships between terms in index entries and other (unstated) terms which might also occur to the user who regard a particular entry term as falling within the scope of his enquiry, including the names of its broader classes and other associated terms. These extra terms are usually excluded from index entries, but are nevertheless present by implication. They are handled in PRECIS by *See* and *See also* references extracted from a machine-held thesaurus.
This exclusion of semantically related terms from subject statements is also, of course, a feature of natural language. Whenever someone speaks of 'computer systems' he does not need to state explicitly that he means tool for 'data processing', nor does he need to refer to 'reprography' whenever one speaks of 'xerography'. In each case, one of the terms (i.e., 'data processing' and 'reprography') is present by implication as part of one's normal frames of reference. Insofar as the present study is concerned, the emphasis will be on the syntactical side of the system, and will touch only briefly on semantic or thesaural relationships. This does not mean that semantic relationships are less important than the syntactic ones, but for the indexer at present "syntax... offers greater challenge" (Austin, 1976a, p. 4).

3.3 Syntax

Syntax can be considered from two viewpoints:

(1) the format and structure of index entries.

(2) the grammar, based on a schema of codes which represent grammatical roles, used to regulate the writing of input strings.

3.3.1 Entry Format

PRECIS differs from many other indexing systems in its extensive and deliberate use of a two-line entry format, as opposed to the one-line entry which is generally found in, for example, subject headings, the chain index, and systems such as KWIC. An entry in PRECIS may occupy two or more lines on the page, and terms can be assigned to any of three basic positions in the entry. This can best be demonstrated by the following example of a string of four terms summarizing a subject.

Great Britian - Public libraries - Chartered librarians - Training
The above name of string ensures that terms in the string are arranged in such a way that:

(a) they form what is called a 'context-dependent sequence, which simply means that each term sets the next term into its obvious context. The concept, 'Great Britain', for example, establishes the environment in which the 'Public libraries' (and the rest of the concepts) were considered; the 'Public libraries' identifies the contextual whole of which 'Chartered librarians' form a part, while the act of 'Training' was applied to the entity 'Chartered librarians'.

(b) they also form a 'one-to-one' related sequence. This arises directly from context-dependency and simply means that each term is directly related to the previous and following terms in the string.

Both of these notions, i.e., context-dependency and one-to-one relations, are recognizable features of natural language. In ordinary conversation, we tend to receive each word of a speech in the context of those which preceded it, and to fully comprehend it frequently we have to wait till the final word is spoken. Similarly, both are important factors in rendering an index entry intelligible and unambiguous, regardless of the index systems in use. In expressing 'meaning' in an index entry, it is possible that one-to-one relations play the more important role. The binding strength of these relationships is visible even when the terms in the string are set down in reverse order:

Training - Chartered librarians - Public libraries - Great Britain

Apart from 'being co-extensive with the subject as well as forming a meaningful statement, it satisfies many of the criteria considered in section 3.1 above.' But problems arise when one considers an entry under one of the middle terms in the string, e.g., 'Chartered librarians'. It would be quite a simple matter to transpose this term to the start of the sequence so
that it would function as an entry point to the index:

Chartered librarians/Great Britain - Public libraries - Training

But, an element of ambiguity has entered into the entry. It is now difficult to tell with certainty, whether the 'Chartered librarians' are being given training, or they are involved in the training of some other persons. The reason behind this ambiguity is that the original one-to-one relationship has been disturbed; automatically, the mind then imposed a new set of relationship, quite different from the original one, so that the role of 'Chartered librarians' is no longer obvious.

For various reasons, however, it was decided that the solutions offered by the keyword systems are not suited to the production of an index to a multi-disciplinary bibliography, such as BNB, where a high level of collocation is desirable. A two-line three-position entry format was proposed, instead of the single-line entry, as a mechanism to preserve the one-to-one relationships. The impetus came from Farradane's work with two-dimensional structures in his system of relational analysis. The outcome is a typical PRECIS entry with the parts separately named as follows:

Heading

<table>
<thead>
<tr>
<th>Lead</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.1: PRECIS Entry Format

The 'lead' is the term which functions as the user's access point to the index and the 'qualifier' contains those terms which establish the wider context in which the lead was considered, while those in the 'display' are context-dependent to the lead. The lead and the qualifier together form the 'heading'. 
3.3.1.1 Standard Entry Format

The adoption of the above structure, with its three named positions not only allowed to display the one-to-one relationship between terms in index entries, it also provided the basis for a fairly simple mechanism for generating a set of entries from a single coded input string. The procedure leads to entries in what is known as the 'standard format' and the operation leading to the generation of entries in this format is known as 'shunting'. At the beginning of the procedure all the terms in the string, organized as a context-dependent sequence, have been arranged in the display position, i.e., the lead has not yet been occupied:

```
<table>
<thead>
<tr>
<th>Great Britain</th>
<th>Public libraries</th>
<th>Chartered librarians</th>
<th>Training</th>
</tr>
</thead>
</table>
```

Figure 3.2: Order of Terms before Shunting

As soon as shunting commences the first term in the string is shunted out of the display and into the lead, and the rest of the terms are shifted to a standard indentation position:

Great Britain

Public libraries. Chartered librarians. Training

When the second entry is generated, the term in the lead is shunted across into the qualifier, and the next term from the display is shifted into the lead:

Public libraries. Great Britain

Chartered librarians. Training

The operation is repeated twice more, giving us the following entries:

Chartered librarians. Public libraries. Great Britain

Training
It can be seen that an elementary transformation occurred during the construction of these entries. Terms which originally appeared in the display in the order A - B - C - D were shunted, as they passed through the lead and into the qualifier, to the reverse of the order, i.e., D - C - B - A. Since the terms in the input string were organized on a context-dependency basis, this ensures that the user who enters the index under a given lead can see, on the same line, the successively wider contexts (if any) in which the chosen term was considered by the author. Other types of transformation, representing two further entry formats, are considered later.

3.3.2 Compound Terms -- Technique of Differencing

The shunting technique described above was intended to deal with a certain category of compound subjects -- that is, subjects which are expressed as more than one term. It was found that a basically similar technique could also be applied to the compound term, i.e., a term, such as 'adult vocational education', which consists of more than one word. Compound terms have been recognized as a source of problem since the days of Cutter, who prescribed the 'significant word' formula to solve the problem. He stated that an entry should be made under the first word in a compound term, except when "...some other word is decidedly more significant, or is often used with the same meaning as the whole name" (Cutter, 1935). In that case the term should be inverted, producing such entries as 'Libraries, Private'. Based on this rule traditional subject heading lists began to invert some multiword headings (e.g., 'Libraries, University and College') artificially, but not always (e.g., 'School libraries'), which led to difficulties and inconsistencies. Post-coordinate systems tried to solve this by factoring such terms into single-word terms (e.g., 'School' and 'Libraries'), but
leading into further problems of loss of original meaning (as in the factoring of 'Professional education' into 'Profession' and 'Education').

In contrast, "Terms in PRECIS are always printed in their natural language order -- there are no inverted headings" (Austin, 1984a, p. 46) [2]. Access can nevertheless be provided under any of the words in a compound term without losing or distorting the meaning of the whole term. This is accomplished by the technique called 'differencing'. A compound term usually consists of a 'focus' and one or more 'differences'. Focus consists of a noun or substantive element which indicates the general class of things, properties or phenomena to which the term as a whole refers, e.g., books, libraries, etc. On the other hand, difference indicates the characteristics which defines a subclass of the focus, by making it more specific, e.g., 'Illustrated books', 'National libraries', etc. Various kinds of differences were recognized in PRECIS:

1) those representing subclass in the form of
   a) an adjectival phrase preceding the focus, e.g.
      \textit{international} networks, \textit{in-service} training
   b) an adjective or adjectival phrase following the focus, e.g.
      public libraries \textit{serving children}, solicitor \textit{general}
   c) a prepositional phrase following the focus, e.g.
      books \textit{in Punjabi}, pamphlets \textit{with Nigerian imprints}

2) dates, e.g.
   \texttt{1900-1950, ca 1800}

3) place treated as an adjective, e.g.
   \textit{Indian} \texttt{manuscripts}, \textit{Eastern Europe}

4) parenthetical expressions used to indicate the names of tests and such.

Out of these, two main structural types of differences, each with subdivisions, have been distinguished in PRECIS:
(a) *Preceding differences*, where the differences precedes its focus, either as a separate adjective (as in 'Compact discs') or as the component of a concatenated word (as in 'Videodiscs').

(b) *Following differences*, where the difference is printed after its focus, either as an adjective (as in 'Attorney general'), or as a noun or nominal phrase following a preposition (as in 'Economies with uncertainties').

3.3.2.1 Preceding Differences

This is one of the areas where PRECIS has developed since the publication of the first edition of the *Manual*. It was realized that the original system which was devised for dealing with typical combinations of nouns and adjectives in English, was not sufficiently flexible or powerful to handle all compounds in other languages, such as German. The new set of codes to deal with these languages has been specified and subsequently incorporated into the second edition of the *Manual*. The new differencing codes are prefixed by a dollar sign; and have two further characters. The first controls whether or not the text following the code is to appear in the lead, and if it should have a space before it when placed before another word. It is summarized in this table and used to generate a set of entries from an input string:

<table>
<thead>
<tr>
<th>Space generating</th>
<th>Close-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead</td>
<td>0</td>
</tr>
<tr>
<td>Non-lead</td>
<td>2</td>
</tr>
</tbody>
</table>

*String:* *(1) industries $21 chemical $12 bio

[Note: Throughout this chapter an asterisk ("**") would indicate a lead.]

*Entries:* Industries

*Biochemical industries*
The second number prefixed to the differences indicates the 'level' of a difference, or its semantic difference from the focus. When a term contains two or more differences at the same level, they are written in the order that produces the most natural output (for a general observation, see Section 3.3.2.4 below), clearly shown in the figure below:

```
 readers    young    handicapped    mentally
       ↓         ↓           ↓               ↓
        focus   1st level 1st level 2nd level
         space-generating space-generating space-generating
         difference   difference   difference
```

Figure 3.3: Levels of Preceding Differences

This will be coded as follows:

*(1) readers $21 young $21 handicapped $22 mentally

and will generate the following entries:

Readers
   Mentally handicapped young readers

Young readers
   Mentally handicapped young readers

Handicapped readers
   Mentally handicapped young readers
   Mentally handicapped readers
   Mentally handicapped young readers

It should be stressed that these differencing procedures are not applied indiscriminately to any compound term, but only to terms which
satisfy certain criteria. These are set down in the form of four basic 'Rules of differencing'. The rules were devised for two reasons: firstly, to avoid over-complex terms in strings and entries; secondly, to ensure consistency among the different members of a team of indexers. Three of these rules are regarded as strong recommendations rather than binding instructions. The first rule, "Do not difference a property or part by the entity or action which possesses it", however, is mandatory. For example, a multi-word term 'computer software' should be factored into separate elements, and reexpressed as two nouns:

*(1) computers
*(p) software

rather than as *(1) software $21 computer

3.3.2.2 Following Differences

In English, as in many other languages, compound concepts are sometimes, expressed in the form of prepositional phrases, such as 'Pamphlets with Nigerian imprints'. This phrase, like the adjective-noun combinations discussed earlier, can be analyzed into a focus and difference, e.g.

\[
\begin{array}{c}
\text{pamphlets} \\
\text{focus}
\end{array}
\quad \downarrow \quad
\begin{array}{c}
\text{with Nigerian imprints} \\
\text{difference specifying a}
\end{array}
\quad \downarrow
\begin{array}{c}
\text{class of pamphlets}
\end{array}
\]

Figure 3.4: Level of Following Difference

In a limited number of cases the indexer can apply the technique described above to terms such as above:

Subject: (1) libraries

(p) stock

*(q) Nigerian imprints $21 pamphlets with

Pamphlets with Nigerian imprints

Pamphlets with Nigerian imprints. Stock. Libraries

However, caution must be exercised in the application of differencing procedure to such terms. It should be avoided unless both the focus and difference share the same context in any circumstances. Failure to do so would result into entries such as the following:

Subject: Pamphlets on dieting in libraries

String: (1) libraries

(p) stock

*(q) dieting $\text{21 pamphlets on}$

Entries: Dieting. Stock. Libraries

Pamphlets on dieting

Pamphlets on dieting. Stock. Libraries

To avoid such misleading constructions (as in the first entry), a new primary code 'h' was suggested to label differencing nouns or noun phrases (Austin, 1982b, p. 230-236). The application is as follows:

String: (1) libraries

(p) stock

*(q) pamphlets $\text{v with}$

*(h) imprints $\text{21 Nigerian}$

Entries: Pamphlets. Stock. Libraries

With Nigerian imprints

Imprints. Stock. Libraries

Pamphlets with Nigerian imprints

Nigerian imprints. Stock. Libraries

Pamphlets with Nigerian imprints

The above procedure even allows to factor the compound noun phrase 'Nigerian imprints'. It collocates entries better by bringing together names of species consistently under the name of the genus, e.g.
Pamphlets
  On dieting
  Publicity pamphlets
  With Nigerian imprints
  etc.

But the application of 'h' must satisfy the following two pre-conditions:
(1) The focus and difference must equally share a one-to-one relationship with earlier term(s) in the string.
(2) The concepts expressed by the focus and difference should be co-located in space and/or time.

Similarly, the noun-adjective constructions proved difficult to conform to standard differencing procedure. Such constructions are frequent in Romance languages, such as French. Several new codes were suggested by French-speaking indexers which would produce entries under both parts of the compound term, and take care of the problem of number and gender of adjectives in these languages. It would ensure that various forms of adjective are all converted, via a lexicon, into the same noun form. But, due to difficulties of generalization and software generation, these codes were finally rejected in favour of a much simpler procedure [3]. It is recommended that terms containing following differences (both prepositional noun phrase and noun-adjective constructions) should be input in their natural language order and access should be provided to the differencing term(s) through See also references extracted from the thesaurus (though with a certain amount of disenchantment from the French-speaking indexers).

For example,

String: *(1) pamphlets on dieting
Entry: Pamphlets on dieting
and this could be supported by a reference such as:

Dieting

See also

Pamphlets on dieting

3.3.2.3 Time as a Difference

Time can be seen from two different viewpoints. In the sense of tense (past, present or future), the concept of time falls outside the scope of syntactic relationships, but specific terms within a statement, referring to time, can be analyzed in terms of deep cases. But linguists seem to have taken less interest about 'time' than their counterparts in documentation science. Evidence is Dewey's introduction of both 'time' and 'space' as common subdivision '-09' and Ranganathan's recognition of 'Time' as a fundamental category, albeit with least significance. In PRECIS, 'time' indicating terms are regarded as differences, on the grounds that they narrow the interpretation of the concept or concepts involved. The numerical forms are preferred, e.g.

String: (1) books
*(2) printing $d 1900-1950

Entries: Printing. Books

1900-1950

3.3.2.4 Universality of Differences

There is no doubting the fact that the differencing procedure discussed above, is especially suited to the indexing of English-type adjectival constructions (e.g., 'skilled personnel'), but the formal analysis on which it is based can be applied with equal effect to similar constructions in other languages (for detailed discussion, see Austin, 1982b, Chap. 10), and also to prepositional phrases (e.g., 'personnel with skills') with certain
limitations. These different constructions are all considered as indicators of species, and they can all be analyzed into some basic components, i.e., focus (such as 'personnel'), plus a differencing concept (e.g., 'skill', which may be expressed as 'skilled', 'with skills', etc.). The focus, always a noun, indicates the general class to which the term as a whole refers. The difference, however expressed, refers to an attribute which, added to the focus, specifies a subclass of the focal concept. A subclass occasionally takes more than one structural form, which is likely to vary from one language to the other (the following example has been quoted from, Austin, 1982b, p. 173):

**English:** Psychology of children  
Child psychology  
Children's psychology

**German:** Psychologie des Kindes  
Kinderpsychologie  
Des Kindes Psychologie

Some linguists may not agree that these phrases convey same shades of meaning or emphasis, but from an indexer's point of view, these are all semantic equivalents. At this point, Austin has agreed that, "differencing calls for a frankly classificatory approach" to the analysis of compound terms, where they could be regarded as variant expressions of a common relationship between the class and its subclasses. However, in modern grammar these are characterised as the 'premodification' and 'post-modification' of the noun phrase, which are conceptually not different from each other (Quirk et al, 1985, p. 1239).

Word order in English adjectival compounds is quite clear when the noun (focus) is modified by preceding adjectives (differences) which can be named only as first and second level differences, and the natural left-to-right order can be applied satisfactorily, e.g.,
Figure 3.5: Word Order in English Adjectival Compounds (Type I)

Here each word modifies the next component of the phrase. However, changing the order of adjectives either makes it a totally different concept or tends to take it beyond normal comprehension (as in 'handicapped visually readers'). Problems arise only when a noun is modified by two or more differences on the same level; for example, when each of two differences is related directly to the same focus, neither difference modifying the other:

![Diagram of word order in Type I compounds]

Figure 3.6: Word Order in English Adjectival Compounds (Type II)

However, in some cases, this order may not be so obvious, especially in the newly-born concepts. During his research with PRECIS, Austin realized that "the order of words in English adjectival compounds is more than a matter of chance, and cannot be explained only in terms of speech conventions. Instead, this order appears to be subject to rules which can themselves be related to an internal categorical system". Such a system of categories was established for determining the preferred order of components in an adjectival compound in PRECIS (Austin, 1984, pp. 309-310).

In section 3.2, we have decided to discuss PRECIS in terms of two separate categories of relationships, viz., syntactic and semantic relationships, but in reality, these two categories have overlapped to some extent, especially in this section. Same is true of any discussion of most language systems. In PRECIS, the manipulation of strings into entries
appears to be entirely syntactical, but semantic relations are considered as well when we deal with compound terms. For example, terms such as the following can be organized as a hierarchy in thesauri as well as into focus-difference relationships as in a compound term. The former approach would generate entries, such as:

Information services

See also

Bibliographic information services

Fee-based information services

Online information services

-- while the latter approach would produce entries, such as:

Information services

Bibliographic information services

Fee-based information services

Online information services

etc.

The indexer's choice of thesaural approach is recommended, if (1) it is a case of following difference, and/or (2) differencing would lead to an excessive number of displays under a given heading.

3.3.3 Schema of Operators

In a PRECIS string, as seen earlier, terms are organized into a sequence according to the principle of context-dependency, and the terms which represent the 'core' of a subject generally from one-to-one sequences. As a general rule, it is a simple matter to recognize when a string of terms has been organized in this way. But for a practicing indexer, this might not be enough; something more definite is needed -- that is, a kind of indexing grammar, which guides the indexer in writing strings in a consistent way. In other words, we need a syntax, a normalizing device, which functions upon
the index system. This is provided in PRECIS by the schema of role operators listed in Appendix 3.1. The basis for the syntax of PRECIS has changed with the development of the system, beginning with the classificatory principles but steadily moving towards an explicit linguistic explanation. Two separate but related schools of thought in linguistics provided the answer: transformational-generative grammar and case grammar.

3.3.3.1 Contributions of Transformational-Generative Grammar

It is generally said that the order of terms in a PRECIS string tends to favour the passive mood as seen in a declarative sentence, on the grounds that the object on which an action has been performed is stated before the name of the action, and the action precedes the name of the agent (Austin, 1976a, pp. 41-42). We have already noted the essence of transformational-generative grammar (Chap. 2, Section 2.3.1.1), that at the base of the surface structure of all human languages lies a deep structure. Humans are capable of generating numerous utterances by applying a schema of rules, called transformations, on this basic structure. For example, all the following constructions convey the same basic thought:

(a) John loves Mary
(b) John is in love with Mary
(c) Mary is loved by John
(d) Mary has John's love

and so on. It is reasonable to suppose that these transformational mechanisms function in both directions, that is to say, they operate during the generation of new utterances, and they also serve as routes between the utterances we hear and the deep structure which contains their meanings. It has been noted that among these the active-passive transformation, i.e.,
John loves Mary
Mary is loved by John

Figure 3.7: Active-Passive Mood Correspondence in English

resembles not only the way a trained PRECIS indexer can visualize the full set of entries which will be generated from an input string during the act of writing, but can also logically infer the original input string quite readily from the set of entries produced by the computer. The principal question which then concerns documentalists is: What form does deep structure take? From their point of view this is a matter of practical concern, since they have to devise a general order of concepts, depending mostly on the deep structure of the message itself. Other natural language functional adjuncts, such as prepositions, conjunctions, etc., he uses very sparingly. It appears that all sentences can be reduced to a set of basic components linked by one-to-one relationships, and generally taking the form:

Noun phrase -- Verb phrase -- Noun phrase
(The librarian) (believes that) (the student borrowed the book)

In the first analysis of a complex sentence as this, each of the noun phrases may itself be a complex system which requires a further stage of analysis, the constituent phrases being themselves expressed in the form shown above. For example,

Noun phrase -- Verb phrase -- Noun phrase
(The student) (borrowed) (the book)

It can be observed from the structure of the sentence as a whole as well as its constituent phrase that both are active voice constructions. Whenever a sentence involves both a subject and an object, transformational-generative grammarians seem to prefer the basic phrase structure in its active form, that is subject-verb-object, while the passive is referred to as the product
of transformation. This favouring of the active rather than the passive must be seen in direct contrast to the approach found characteristically in controlled information languages, where the order most commonly used, tends to favour the passive mood, that is the object-verb-subject. As far as the 'core' concepts based on the action are concerned, Vickery's standard order, which prescribes the citing of the patient before the operation before the agent, is similar to the order of words in the passive declarative sentence which forms the basis of input strings in PRECIS. The same passive order is implicit in Coates' (1968) rules for BTI: "A term denoting an end product takes precedence over a term denoting an activity leading up to or acting upon the end product, and if there is an agent concept involved, a term for the agent follows that of the activity". It is true that the active constructions in English are usually shorter, and appear to be simpler and more direct, than their passive counterparts. But the assumption that deep structures should be represented by active constructions is not unquestionable (Austin, 1982b, pp. 149-150).

However, the final outcome of this active-passive discourse and evidence offered by the transformational-generative grammar is, that we have a natural tendency, when designing ILs, to categorize concepts into particular types: and in those systems which involve pre-coordination, we are also predisposed to cite these components in a certain order. It also contributes to the fact that, as human beings, we have been programmed to organize our thoughts in a certain way.

3.3.3.2 Evidence from Case Grammar

At this point of discussion, one might began to have the feeling that, the explanation offered in the previous section is enough to support a linguistic basis of PRECIS. But, it is not. Austin soon realized that, Chomsky's deep structure and PRECIS' adoption of passive declarative
sentence, were not enough to explain every possible string that might be needed to provide a meaningful logical syntax of terms within a set of index entries. In certain cases, logical explanations called for reference to relationships that lie outside the descriptive range of transformation-generative grammar. Ambiguity arises whenever a given surface structure, e.g., 'The killing of the tigers was sporadic', can be mapped on to more than one deep structure, where the term 'Tigers' functions as the object in one of the kernel sentences, and as the subject in the other. Such ambiguities can only be solved by distinguishing action concepts through reference to the kinds of noun with which the action can be associated, and the nouns themselves were specified in terms of their logical roles, such as 'object', 'performer' or 'possessor'. This theory has been proposed by workers such as Fillmore, who pointed out that transformations cannot be explained adequately without reference to inter-concept relationships more specific than those between noun phrases and verb phrases; a more revealing analysis of deep structure would be in terms of 'cases', such as agentive, instrumental and dative, since these give clear indication of word functions in relation to each other. One interesting thing to note is that, it is possible to map components of deep structure phrases on to surface structures regardless of how they are expressed. For instance, in the passive construction, 'Mary is loved by John', the agentive deep case would still logically apply to 'John'. To some extent, a similar situation was described by Austin in the first edition of Manual:

"Whenever reference is made to the object or performer of an action, this invariably means the logical, not the grammatical, object or performer. These are not necessarily the same thing".

That is to say, the indexer has to look somewhat below the surface structures. Fillmore introduced the term 'deep case' as a generic to cover a wide range of syntactical situations and offered a basic set of deep cases
as linguistic universals. Languages vary in the means they employ to convey deep case relationships in surface structures, i.e., they differ in their use of 'case markers', which may be either inflections or prepositions or word order or any combinations of these three. Apart from Fillmore, authors such as Chafe, Grimes, Longacre, Johansen, etc., have also considered deep cases and their categorization. From an indexing point of view and certainly that of PRECIS, the following are considered important and regarded as truly language-independent, and these are placed below side-by-side with their parallel (approximately) operators in PRECIS:

<table>
<thead>
<tr>
<th>Deep cases</th>
<th>PRECIS operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Agentive:</td>
<td>Peformer of intransitive action (key system = 1)</td>
</tr>
<tr>
<td></td>
<td>Performer of transitive action (agent = 3)</td>
</tr>
<tr>
<td>(b) Instrumental:</td>
<td>Performer of transitive action (instrument = 3)</td>
</tr>
<tr>
<td>(c) Experiencer:</td>
<td>An animate entity registering the action, but without undergoing a change in state (key system = 1)</td>
</tr>
<tr>
<td>(d) Factitive:</td>
<td>Thing that completes the action; the product of the action (key system = 1)</td>
</tr>
<tr>
<td>(e) Locative:</td>
<td>A spatial constraint on the topic (location = 0)</td>
</tr>
<tr>
<td>(f) Objective:</td>
<td>Object of transitive action (key system = 1)</td>
</tr>
<tr>
<td>(g) Benefactive:</td>
<td>Entity towards which an action is directed (key system = 1)</td>
</tr>
</tbody>
</table>

3.3.3.3 Deep Cases and Role Operators

Any one familiar with PRECIS can already deduce certain similarities between the deep cases mentioned above and PRECIS' schema of role operators (see Appendix 3.1), e.g., between 'agentive' and '3 Agent', 'objective' and '1 Object', and so on. In some situations, however, this apparent consistency would break down, for example, an 'Agent' is usually coded '3', but sometimes '1'. A closer match between cases and operators can be achieved
only if the operators are matched not only against the roles that entities can assume in a subject but also the kinds of action with which they are associated. The deep cases only apply to nouns. Verbs do not take cases; instead their 'case frames' determine the kinds of noun that can be present in a sentence, and the roles that these nouns can assume. Verbs are classified by linguists into: states, processes, actions and action/processes. Austin (1982b, pp. 158-166) has presented a case-by-case survey, where deep cases are related to the kinds of verb with which they can occur, with suitable examples from sample PRECIS strings. However, within the scope of this study it is not possible to elaborate them. The following table (adapted from Austin, 1982b, p. 162) shows the correspondence between the deep case 'Beneficiary' (or 'Benefactive'), representing an entity indirectly affected by actions involving transfer (such as 'give', 'teach', etc.) and PRECIS role operator '1 Object':

<table>
<thead>
<tr>
<th>Class of verb</th>
<th>Subject of statement</th>
<th>PRECIS string</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>Students' attitudes</td>
<td>(1) students</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(p) attitudes</td>
</tr>
<tr>
<td>Process</td>
<td>Theft of audiocassettes</td>
<td>(1) libraries</td>
</tr>
<tr>
<td></td>
<td>from libraries</td>
<td>(p) stock</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(g) audiocassettes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) theft</td>
</tr>
<tr>
<td>Action</td>
<td>Christmas parties for workers</td>
<td>(1) workers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) entertainments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(q) Christmas parties</td>
</tr>
<tr>
<td>Action/Process</td>
<td>Donations to flood victims by students</td>
<td>(1) flood victims</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) donations $v by $w to</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3) students</td>
</tr>
</tbody>
</table>

Table 3.1: Correspondences between the Deep Case 'Beneficiary' and Operator '1 Object'

The role operators in PRECIS have various functions: right from determining the order of terms in the string to the final format of entries and its associated punctuation. From the linguistic point of view, however,
these codes serve two principal functions (Sørensen and Austin, 1976, p. 114):

(a) Compound subjects should be analyzed into conceptual units according to common frames of reference; that is to say, deep cases should be identified.

(b) The built-in filing value of operators ensures that a team of indexers will consistently achieve the same order in input strings; this means, the position of a term in a string (or deep structure) is determined primarily by its role or case.

In view of the two functions listed above and in terms of linguistic analysis, the operators and codes are said to fall into three main groups:

(a) those which identify deep cases: operators 0 to 6, s to u, and code $d;

(b) those indicating semantic relationships between concepts occupying the same deep case: operators p to r, and differencing codes. For convenience, operators f and g, indicating coordinate concepts, could be regarded as a member of this group, although they logically indicate disjunction, similar to the Boolean 'or'.

(c) those which function essentially as case markers: s to u, $v and $w.

It is possible to recognize a direct connection between the terms introduced by primary operators 0 to 3 and certain everyday parts of speech. For instance, the operator 0 is related to the locative case in grammar, while the operators 1, 2 and 3 introduce terms which usually correspond to the object, verb and subject in a sentence. It should be noted that the operators 4 to 6 identify concepts which stand outside the representation of subject semantic content, such as the author's viewpoint, or the physical form of the medium (Hutchins, 1975, p. 105). These constitute essential components of a full subject statement, but since they do not present problems from a truly linguistic viewpoint, their discussion will be limited only to the special entry format they create. The dependent elements, such
as p and q, do not strictly indicate grammatical roles as such. The role of a term which is prefixed by a dependent element code is determined instead by whichever primary operator was assigned to an earlier term in the string, and so sets the dependent element into its logical context. It can also be seen that some of the operators belong to more than one category.

3.3.4 Special Routines with Linguistic Orientation

PRECIS, as all other indexing languages, uses certain special routines to generate meaningful index entries. A couple of such routines deserve special attention as these have linguistic implications. The first is the concept of 'substitute' and the second is the concept of 'theme interlink'.

A 'substitute' can be defined as a noun phrase which being inserted into a string, replaces a set of terms when a term outside the set appears in the lead. Two forms of substitute are available: (a) upward-reading substitute and (b) downward-reading substitute. In both the cases the substitute phrase is preceded by a convention which consists of 'sub', plus a number from 1 to 9 which indicates the number of next higher or lower terms which they are to replace, plus the upward or downward arrow. For instance,

String: *(2) agriculture
(p) research
(sub 2 ↑)(2) agricultural research
*(2) planning

Entries: Agriculture

Research. Planning

Planning. Agricultural research

Substitutes are used in various circumstances, two of which are relevant in the present context:
(i) To resolve ambiguity, as in the second entry above. Without the substitute, this entry would have appeared as below:

Planning. Research. Agriculture

-- where the user could not immediately deduce whether the research was being performed on 'Planning' or on 'Agriculture'. The substitute thus links together the related terms, and introduces a noun phrase which functions as a single semantic unit from the viewpoint of 'Planning'. It also reveals another aspect of PRECIS with linguistic implications. It represents a typical family of strings containing what are known as 'second actions', indicated by the repetition of the operator 2, and as shown in the following string:

*(2) collective bargaining
(s) role $v$ of $w$ in

*(3) trade unions
(sub 3 $\uparrow$)(2) role of trade unions in collective bargaining

*(2) research

This string involves a second action as well, but it follows a sequence of concepts related in a more complex manner. To produce a meaningful entry under the final term 'Research', a substitute phrase is needed in all NL using prepositional constructions (as in English):

Research. Role of trade unions in collective bargaining

In this case a step-by-step analysis of concepts produces a sequence of terms at the beginning of the string which is not amenable to a noun phrase transformation, and a substitute becomes necessary.

The second mechanism requiring an explanation involves the use of the 'theme interlink codes', z, x and y. One of these codes has to be written in a predetermined position in the manipulation code which precedes each term in an input string. The code z identifies a common element, whereas x identifies the start of a separate sub-theme, and y subsequent elements in
each such sub-theme. The typical occasion on which these codes are used, is to allow the indexer to deal with a multi-topic document, such as 'Remuneration of personnel and enhancement of skills of personnel', by using a single string, e.g.,

String: *(z)(1) personnel
*(x)(2) remuneration
*(x)(p) skills
(y)(2) enhancement

Entries: Personnel
Remuneration
Skills. Enhancement
Remuneration. Personnel
Skills. Personnel
Enhancement

But, they are also used to deal with a special class of problems which is more linguistic in character, and which again arises from the fact that a given compound notion needs, for indexing purposes, to be analyzed into separate concepts, yet it constitutes a single semantic unit from the viewpoint of some other term in the string. For instance,

*(x)(1) personnel
*(y)(2) migration
(y)(2) effects of remuneration differentials
*(x)(2) remuneration
*(y)(p) differentials
(y)(2) effects on personnel migration

3.3.5 Stages of Concept Analysis, and the Use of Operators

Sørensen and Austin (1976) have proposed a three-stage model of concept analysis which can explain the production of correctly structured and
logically acceptable PRECIS strings in any NL, including English. The theoretical model of the indexing process is based on the presumption that an indexer has already examined the document and determined its subject content in the form of a 'title-like phrase' which is cast in the active mood, e.g., 'This is a case study on planning the development of public libraries in Great Britain'. This procedure is now set down in a code of practice prepared by the International Organization for Standardization (1982). The above phrase might be regarded as a surface structure statement, and the string derived from it a deep structure source of various index entries. The three stages are:

Stage 1: Identify concepts in terms of their deep cases.

Stage 2: Identify embedded deep cases which may be present in concepts already analyzed at stage 1. This stage is recursive.

Stage 3: Establish the relationships between concepts which occupy the same deep case. This stage is also recursive.

Stage 1 is common to all subjects, simple or complex. This stage along with the stage 2, deal with syntactical relations, while stage 3, involves a variety of semantic relations. For example, a simple subject, such as 'Planning the development of public libraries' would be analyzed at stage 1 as:

(2) development of public libraries and (2) planning

yielding two actions, each occupying a different case as shown by the repeated operator 2. At stage 2, the first phrase, more complex by nature, is further analyzed into an action and its object, as:

(2) development and (1) public libraries

The object term 'Public libraries' is finally subjected to a semantic analysis at stage 3, where its components are identified as a focus and its first level difference (as shown in Section 3.3.2), as:

(1) libraries $21 public

Hence, the final string is:
(1) libraries $21 public
(2) development $w of
(2) planning

It should be noted that, the case marker, '$w of', has been added deliberately to make explicit the relationship between the phrasal components first identified at stage 1.

By combining these three stages of analysis, a model of four subject types was proposed:

Type 1: those requiring analysis only at stage 1, e.g.,
String: *(1) libraries
          (2) management
Entries: Libraries
          Management

Type 2: those calling for analysis at stages 1 and 2, e.g.,
String: *(1) libraries
          (2) management
          (sub 2 ↑)(2) library management
          (2) research

Type 2 subjects require the repetition of at least one core operator, either explicitly or implicitly, but without calling for the use of dependent element codes or differences.

Type 3: those requiring analysis at stages 1 and 3, e.g.,
String: (0) India
        *(p) West Bengal
        *(1) libraries $21 public
        (2) development

Type 3 subjects consist of concepts sharing one-to-one relationships
expressed by primary operators, plus dependent elements and/or differences, but without involving embedded structures.

**Type 4**: those which calls for analysis at all three stages, e.g.

String: *(0) Great Britain (LO)*  
*(p) Northern Ireland  
*(1) personnel $21 skilled  
*(2) migration  
(sub 3 †)(2) migration of skilled personnel in Northern Ireland  
(2) prevention $w$ of  
*(2) research

From a linguistic viewpoint, especially multilingualism, subjects belonging to types 3 and 4 are of special interest. These both involve a stage 3 analysis, and so call for the identification of relationships between concepts occupying the same deep case. That is why, the stage 3 analysis described above is a semantic rather than syntactic operation. Due to this change of nature, and because these operators add certain amount of complexity in the string, it has been decided to discuss them in the following sub-section entitled 'dependent elements'.

3.3.5.1 Dependent Elements

Clearly, at the beginning the reader might feel confused at the mention of the word 'semantic' in describing relationships which are to be handled by role operators rather than assigned to the thesaurus, as mentioned earlier (cf. Section 3.2). The origin of the confusion can be attributed to the change in practice in the handling of certain operators, which are now delegated to represent some semantic relations: in particular, the differencing codes (described in Section 3.3.2) and the dependent element codes p to r. In fact, this is the fuzzy zone between two aspects of
linguistics, viz., syntax and semantics, recognized by Gardin (1973) and others. At the end of section 3.3.2.3 earlier, we have already pointed out this lack of distinction and the alternative solutions available to the indexer for the treatment of compound terms, viz., either the use of thesaural relationships or the rules of differencing. Here we would like to add a third option; the compound concept could be factored into two separate terms, each introduced by its own operator. For the sake of convenience, it is better to demonstrate them together by considering all the three ways in which a term such as 'Women personnel' can be handled:

(a) Use of thesaurus:

Personnel

see also

Women personnel

(b) Using differencing procedures:

String: *(1) personnel $21 women

Entries: Personnel

Women personnel

Women personnel

(c) Using dependent elements:

String: *(1) personnel

*(q) women

Entries: Personnel

Women

Women. Personnel

Neither linguistic nor logical grounds are decisive in suggesting one option's superiority over the other two. Here again, the choice is determined by pragmatic factors, such as the number of displays permitted under a lead term, need of the special subject field and nature of the collection to be indexed. Observed from a linguistic viewpoint, a similar
situation exists in the case of compound terms which contain the names of their parts or materials, e.g., 'Computer software' (a 'part' defined by its containing whole) and 'Steel racks' (a 'whole' defined by its material). Concepts belonging to the first category can be handled by either method (a) or method (b) mentioned above, but not method (c), because it would express a completely different subject, i.e., steel which can be used in any kind of racks. On entirely theoretical grounds, whether linguistic or logical, it is possible to apply the same reasoning in the other category of concepts also, i.e., 'Computer software'. Apart from being natural and forming a single entity, both parts of the compound term occupy the same deep case. But the rules of PRECIS, especially the differencing rules, make it mandatory to factor such terms into 'Computers' and 'Software', e.g.:

*(1) computers
*(p) software

Here again, rules such as this were formulated for entirely pragmatic reasons which apply to an IL such as PRECIS, but not necessarily to NL. Extra-linguistic explanations for inclusion of concepts marked p and q, are that they clearly raise both the level of exhaustivity and specificity of the string, e.g.

*(0) Great Britain
*(1) engineering industries
*(p) skilled personnel
*(q) women
*(2) training $21 in-service

It can be seen that the inclusion of the term 'Women' as a special class of 'Skilled personnel' has increased the specificity of the input string. The differencing codes also serve similar purposes, as can be seen from the use addition of 'In-service' to more general concept 'Training'. However, this subject still basically has the same 0-1-2 string, and straightforward
shunting would generate entries in the standard format.

Many documents deal with concepts which share mutual or coordinate relationship with some other component in the input string. This coordinate concept calls for addition of a second dimension to the linear or one-to-one structure among concepts in the input string. For example, a subject like 'Selection and acquisition of serials in libraries' could be diagrammatically represented as,

```
Libraries ---- Serials
```

--- Coordinate block

```
Selection

Acquisition
```

**Figure 3.8: The Structure of Coordinate Relationship**

This structure is represented in the following string:

*(1) libraries
 *(p) stock
 *(q) serials
 *(2) selection $v \&$
 *(g) acquisition

--- Coordinate block

-- which will generate the entries as follows:

Libraries

Stock: Serials. Selection & acquisition

Serials. Stock. Libraries

Selection & acquisition


Acquisition. Serials. Stock. Libraries

According to some linguists, sentences containing coordinate concepts belong to a general class of compound statements (but, differ from the compound subject, in its usual sense) and generally occupy the same deep
case provided they satisfy the criteria that 'only noun phrases representing the same case may be conjoined'. Among specific varieties of coordination, 'succession' and 'coupling' have parallels in indexing, which is matched by the use of operator $g$ in PRECIS. 'Succession' refers to the deep structure of concepts, which is linked in a coordinate fashion forming a 'time-dependent sequence'. The order of terms in the coordinate block in the above string follows this sequence, which normally follows an alphabetical order. The latter type is most often experienced in indexing and would be classed as 'coupling' following the linguistic terminology. Most topics containing coordinate terms are conceptually similar to the examples seen above, and could be expressed without loss of meaning by writing separate strings, such as:

*(1) schools and *(1) schools
*(p) teachers *(p) teachers
*(p) administrative skills *(p) teaching skills

These will generate entries under 'Administrative skills' and 'Teaching skills' as:

Administrative skills. Teachers. Schools
Teaching skills. Teachers. Schools
-- which exactly match those produced from the earlier string, since the rest of the members of the coordinate block is automatically deleted from the entry when one of its terms forms the lead. But, occasionally, this may result into an unsatisfactory entry, such as:

*(1) schools
*(p) teachers
*(p) administrative skills $v \&$
*(g) teaching skills
*(2) integration

where the last term 'Integration' acts as a 'binding term' for the whole of the coordinate block. Such a string would produce misleading entries, as
follows:

Administrative skills. Teachers. Schools
Integration
Teaching skills. Teachers. Schools
Integration

This raises an immediate question 'Integration with what?' A new code f has been used to replace g whenever a binding term is present in the string, e.g.,

*(1) schools
*(p) teachers
*(p) administrative skills &
*(f) teaching skills
*(2) integration

-- which displays the whole of the coordinate block once any of its members appear in the lead, as in:

Administrative skills. Teachers. Schools
Administrative skills & teaching skills. Integration

The special nature of these 'binding term' situations has been recognized by linguists and referred to as 'nominal concretions' present in the deep structure of language.

3.3.6 Inverted Format: Extra-core Operators

As indicated in an earlier section, a special entry structure, known as 'Inverted format' is associated with terms which are prefixed by operators in the range 4 to 6 or any of their dependent elements. When a lead is generated under any of these terms, the display consists of the earlier terms in the string selected in their input order. Terms in this group also generate a special typography, viz., italics, when written in the display. For example, a subject such as 'Bibliography on adult education in India'
would be input as:
*(0) India
*(2) adult education
*(6) bibliographies

This would produce three entries:
India
   Adult education -- Bibliographies
Adult education. India
   -- Bibliographies
Bibliographies
   India. Adult education

3.3.7 Entries Generated by the Predicate Transformation

The third and last of the three formats used in PRECIS is generated by a routine known as 'predicate transformation'. This format is produced whenever a string:
(a) contains a term which represents an agent, coded 3, and marked as a lead; and
(b) the agentive term follows immediately after a term which was prefixed by one of the operators 2 or s or t (each of which indicates an action of some sort).

Both of these circumstances occur together in the following string representing the subject 'Damage to libraries by earthquakes in Japan':
*(0) Japan
*(1) libraries
*(2) damage $v$ by $w$ to
*(3) earthquakes

This string would give rise to the following entries:
The above string and entries depict the role played by the two connective
codes, $v$ and $w$, which essentially act as explicit case markers to the deep
case indicating role operators, whose role otherwise would have remained
implicit. The effect of predicate transformation can be seen in the fourth
of the entries above, where the term 'Earthquakes' is in the lead.
Application of a straightforward shunting procedure would have printed the
phrase 'Damage to libraries' in the qualifier position, not in the display,
in this final entry. But marking of 'Earthquakes' as 3, led the action
'Damage' to be printed in the display, carrying with it any phrase
components indicated by the connective $w$. Together, these constitute a kind
of predicate. This procedure accomplishes two objects:
(a) It ensures that the actions in which an entity is engaged are
collocated, together with the names of its parts and properties (and to
some kinds) in the display.
(b) It also ensures that it is possible to retain one-to-one relationships
between concepts which may have become separated in the input strings,
due to their different syntactical roles.
In the example above, 'Libraries' and 'Earthquakes' are both related to the
location 'Japan'. But, a simpler two-term subject 'Earthquakes in Japan',
would be coded as follows:

*(1) Japan

*(2) earthquakes

and this would produce the entries:
The predicate transformation ensures that the entry under 'Earthquakes' produced from the earlier, more complicated string, collocates exactly with that produced from the simpler string seen above, even though the term 'Earthquakes' was treated quite differently and was separated from 'Japan' in the input string. Two other operators, which generate predicate transformation and are also important from a linguistic point of view, are the operators s and t. As these involve a different kind of problem from that of straightforward predicate transformation case, i.e., object-action-agent format, they are dealt with in the next sub-sections.

3.3.7.1 Role Defining Terms and Directional Properties

When analyzing a compound subject, such as 'Damage to libraries by earthquakes', it is fairly easy to deduce that the term 'Earthquakes' indicates the agent which is directly responsible for the 'Damage'. However, in many subjects the link between the agent and the action is rather less obvious; either the agent is responsible for a diffuse range of actions which cannot be specified, or its agentive function may be 'indirect', i.e., it serves as a tool or instrument of some other unspecified agent. In such circumstances it is necessary to insert a term or phrase deliberately into a string to explain the role of the agent. These 'role defining' terms are introduced by the operator s. For example, the subject 'Role of trade unions in the management of banks' would be input as:

*(1) banks

*(2) management $w of

(s) role $v of $w in

*(3) trade unions
where the term 'Role', prefixed by s, has been introduced to explicate the relationship between 'Management' and 'Trade unions'. The entries would be:

Banks
  Management. Role of trade unions
  Management. Banks
  Role of trade unions
Trade unions
  Role in management of banks

The operator s is usually used in this way to introduce terms and connectives, such as 'Role-of-in', 'Applications-of-in', 'Effects-of-in', etc., which have simply an explanatory function, and do not justify marking as leads. The relationship involved is also similar to the instrumental case in grammar. It can be seen, however, that they have an almost 'transitive' quality — that is, they are actively related to some term earlier in the string — but it was later realized that those terms are also invariably 'properties' of the next term later in the string, viz., the agent. This special characteristic is evident in the input string and entries produced from the simple subject 'The role of trade unions', as follows:

*(1) trade unions
*(p) role
Trade unions
  Role
  Role. Trade unions

The use of operator p indicates that the trade unions have a role, just as they have various other attributes. Also in the entries, they have been set down in a different order from that of the earlier string. Terms such as 'Role' in the examples above have been given a special name 'Directional properties', which can also form a lead. A typical example of such terms is
'Attitudes', which forms the part of a person's mental make up and should be coded as p, e.g.,

*(1) students
*(p) attitudes

and would generate the following entries:

Students
Attitudes

Attitudes. Students

But, in the ensuing example this simple property is directed towards something else, viz., 'Coursework', and becomes a 'directional property'. Directional properties, because of their active nature, are coded s, as in:

*(2) coursework
*(s) attitudes $v$ of $w$ to
*(3) students

-- which would give rise to the following entries:

Coursework

Attitudes of students
Attitudes. Students
To coursework

Students

Attitudes to coursework

The entry under 'Attitudes' now serves the twin purpose: it takes account of the nature of the term as a property as well as an action. This feature has been referred to as the 'Janus-faced nature' of operator s, when it deals with 'directional properties'. However, Austin (1976a, p. 24) has failed to provide an adequate explanation for this phenomenon, or need for this format, in linguistic terms. But, later, he conjectured his above admissions on the presumption that, such cases have probably been identified by indexers rather than linguists only because the former have paid particular attention to a highly-stylized subset of utterances where such factors tend
to become important. Moreover, indexers have studied such (and many other) statements from a narrow and severely practical viewpoint (i.e., consistency and normalization) that would hardly interest a linguist (Austin, 1982b, pp. 186-187).

3.3.7.2 Author-attributed Associations

In all the examples considered so far, the actions have belonged to a category which is known in PRECIS as 'system initiated'. This means that the action was carried out, directly or indirectly, by a system which is named in, or implied by, the other elements in the string. This section deals with a different class of actions, known as 'author-attributed associations' coded t. These are actions 'which an author is carrying out'; these describe the author's method of treatment of his or her subject, e.g.,

*(1) subject heading lists
(t) compared with
*(1) thesauri

This operator indicates that the act of comparison was performed by the author, not by either of the entities being compared. Special typographic and format instructions have been built into this operator, which could be seen from the resultant entries:

Subject heading lists

compared with thesauri

Thesauri

compared with subject heading lists

It should be noted that, this is the only case in PRECIS where actions are not expressed as nouns, but participles. The use of italics in the entry (which is used as a de-emphasizing device in PRECIS, as in the case of extra-core operators) also indicates that such actions belong to a category falling outside the core of the subject. Their lack of explicit linguistic
explanations has been attributed to the same factors mentioned at the end of the previous section.

3.3.7.3 Two-way Interactions

Another kind of reciprocal relationship can also be treated in a single input string. This is the case of two-way interactions, but these are system-oriented rather than author-initiated, as we have seen above. Actions belonging to this category can be recognized by certain features when indexing:

(a) they always involve at least two other concepts, one on either side of the interaction; and

(b) these interrelated concepts serve simultaneously as object and performer of the same action.

Earlier versions of PRECIS used operator 2 to represent such interactions, but recently a new code u has been allocated to isolate the two-way interaction, as seen in the example below:

String: *(1) employers
*(u)-cooperation $v with $w with *(1) trade unions

Entries: Employers
Cooperation with trade unions
Cooperation. Employers
With trade unions
Cooperation. Trade unions
With employers
Trade Unions
Cooperation with employers

The coding of the terms on either side of the action coded u, as 1, indicates that both 'Employers' and 'Trade unions', share the role of key
systems and also the role of the performer. On the other hand, the two-wayness of the action term 'Cooperation' is shown by the prepositions attached to it. This very factor justifies its separate treatment from other one-way actions considered earlier. This is again one of those special cases which does not have a parallel explanation in terms of linguistics. This brings us to an end of discussion of PRECIS' schema of operators (excluding 'Term codes' and 'Typographic codes'), which have some linguistic orientation. However, Austin (1982b, p. 130) was always aware of the limitations of such an endeavour:

"The ways in which subject statements (e.g., titles, subject headings and index entries) convey their meanings cannot always be entirely explained through paradigms developed for the analysis of natural languages."

The following sections cover some alternative explanations which were also suggested by Austin and others as forming the basis of PRECIS' claim of universality.

3.4 Need for Alternative Explanations

We have already mentioned Austin's arguments adducing this view in section 2.3.1, chapter 2. Apart from the evidence provided in the previous sections in this chapter, further evidence can be quoted from Austin's thesis. It reported the result of a test which showed that a large majority (85.0%+) of indexers, with various levels of experience and at independent locations, produced similar input strings from the same set of subject statements. This high rate of performance explains that, indexers are aware of deep case relations (at least intuitively) which they apply while indexing, and as these deep cases are considered linguistic universals, the operators which mark these cases in input strings also comprise a language-independent system. But, this is a general deduction, which might prove inadequate on specific occasions. For example, the subject statements and strings used in
a general explanation are excessively simplified, which may not be true in an actual indexing situation, where the indexer faces much more complex subjects (this, of course, applies to the present description as well). Examples of such problematic cases were reported by Mahapatra and Biswas (1985) in an earlier work. Austin also suggested that, although linguistic factors, probably operating at an intuitive level, are likely to influence the design of ILs, we also need to realize that these 'contrived languages' are custom-built for special (non-linguistic) purposes which could take them into a different category from NL, to the point where they need their own rules and explanations. He produced a list of such concept classes and relational situations that are treated as 'special' in descriptions of PRECIS, on the grounds that they call for additional non-standard codes, routines or formats to produce satisfactory index entries. We have already discussed them, but a brief listing may be useful:

(a) the Janus-like nature of directional properties;
(b) the attributes of two-way interactions;
(c) the case for a substitute phrase;
(d) the special case associated with second actions;
(e) categorization of adjectives and their relative positions within a noun phrase; and
(f) distinction between system-oriented actions (operator 2 and s) and author-attributed associations (operator t).

As pointed out at the end of section 3.3.7.1, some of these special situations are exclusive to the field of information science. Their lack of identification by linguists is due to the special nature of utterances used by indexers, which is usually absent in NLs. Factors, which tend to become important in ILs -- that is, statements characterized by only one mood (indicative), only one tense (the vague present), and no concern with person (i.e., 1st, 2nd or 3rd person) -- form a highly-formalized subset of
utterances, which is probably too restrictive from a linguist's point of view. A couple of non-linguistic explanations have been put forward by Austin and others as alternatives to the principles mentioned in previous sections: (i) time as an organizing principle; and (ii) presence of a set of basic relationships.

3.4.1 Time as an Organizing Principle

This principle, originally proposed by Sørensen, and first discussed by Sørensen and Austin (1976) in the context of PRECIS, was later realized to have possible wider application as a general organizing principle in all ILs. This principle was stated as follows: "The terms in a PRECIS string are organised as a sequence according to their relative time of conceptualization as determined by their rules." In simple terms it means, a thing (say, 'Wheels') cannot be labelled as a part except through reference to a prior conception of this whole (say, 'Cars'). However, it was emphasized that time of conceptualization does not correspond, except by coincidence, to the order in which things or phenomena occur in real time.

The difference can be illustrated with the following example of a subject containing a second action, such as 'Planning the automation of libraries'. In terms of real time the act of planning precedes the act of automation, which even may or may not take place in future. But in terms of time of conceptualization, this order is reversed, because 'Planning' itself relates to the object phrase 'Automation of libraries', which can be further analyzed as an action, 'Automation', and its pre-conceived object, 'Libraries'. Hence, the order of terms in the string:

(1) libraries
(2) automation $w$ of
(2) planning

-- represents the order in which concepts were conceptualized. In some
respects it is analogous to the difference noted earlier between grammatical and logical objects and agents of actions. As a recapitulation, we can mention the order of terms in the coordinate block, which was also arranged by following this principle. Originally, this principle was thought as a reformulation of the principle of context-dependency, perhaps also amounting to a reformulation of Ranganathan's Wall-Picture Principle. But later it was found to be capable of explaining other principles, such as passive voice order, 'standard' paradigmatic relationships, etc., used to organize terms into meaningful sequences in majority of ILs.

3.4.2 The Primitive Relationships

Another possible corollary following from the context-dependency principle is that pairs of terms which have the 'strongest' links between them in a description should be adjacent in the citation order. In an attempt to establish, categorize and order all such links that may exist between term-pairs, Austin (1982a) proposed a hypothetical set of basic relationships, which will be able to offer an alternative explanation to the various roles indicated by the PRECIS operators. The following two main classes of relations between concepts have been established empirically and appear to be common to any kinds of subject statements, including indicative sentences in NL, terms organized as networks in thesauri and classification schemes, order of terms in pre-coordinate indexes, etc.:

Class 1: Grammatical relations
- Predicative (to be)
- Possessive (to have)
- Active (to do)
- Locative (to locate)

Class 2: Logical relations
- Coordination (and)
- Disjunction (or)
- Negation (not)

The grammatical relations are said to be different from those syntactical relations studied by linguists. They are also different from logical
relations comprising Class 2, in the sense that they cannot link concepts into fully formed syntagms, to the extent that 'A is B' is complete, but 'A and B' is not although it can function as parts of complete sentences. Logical relations are characteristically Boolean functions used in on-line searching. Only one of these is represented in PRECIS, the 'and' relationship, which links coordinate concepts (through f and g) as well as coordinate themes (code x an y). With the exception of 'to do' relationship, all other basic relations are capable of linking only pairs of concepts, one on each side of the relation; for example, 'to be' can link (a) a broader term and a narrower term; (b) a pair of synonymous terms, etc. The same is true of the 'to have', 'to locate' and logical 'and' relationships. The 'to do' relationship can interrelate a varying number of concepts depending upon the class to which the action itself belongs. For example, an intransitive action can link only one concept (as 'Children sleep'), whereas a transitive action may link two or more concepts, each possessing a different role (as 'John broke the window with a hammer'). Other kinds of action need different operators (s, t, u) and generate special formats. Following example introduces all the five kinds of relationship listed above:

---

(0) Calcutta

----

-LOCATE

------

---DO

-----

BE

(1) children $21 destitute

(2) education $21 primary $32 pre-

(3) parents $v &

(4) social workers

---

Figure 3.9: Display of the Primitive Relationships (String)
The same relationships are carried automatically into the entries, of which only the first has been shown below:

```
----------Calcutta
          DO
           BE
           BE
           BE
Destitute children. Pre-primary education.

--LOCATE--
          DO          AND
Attitudes of parents & social workers

----------HAVE------
```

Figure 3.10: Display of the Primitive Relationships (Entry)

A related idea, suggested by Austin, is that searchers normally interpret adjacent terms as being linked in the strongest way possible. He categorized the strength of linktypes with the formula:

\[ \text{Be} > \text{Have} > \text{Do} > \text{Locate} \]

which expresses the notion that 'BE is stronger than HAVE, HAVE is stronger than DO, and DO is stronger than LOCATE'. According to this formula, 'Women. Teachers' is more likely to be interpreted as 'Teachers who are women' ('to be') than 'Teachers who teach women' ('to do'); and 'Teaching. Costs', as 'Costs of teaching' ('to have') than 'Teaching the subject of costs' ('to do').

It has been observed that these primitive relations can be detected not only in the relation between concepts in strings and entries (syntactic level), but also in references extracted from the thesaurus (semantic level). For example,

(a) 'to be' relationship links a genus and its species, e.g.:

- Animals
  - Vertebrates
    - Birds

-- also it links a compound term and its focus, e.g.,
Education

Secondary education

(b) 'to have' relationship links whole to their parts and properties, e.g.,

Trees:

Fruits

According to Austin, an awareness of the primitive relations offers a simple check on the logical adequacy of any system. They are also important from a pedagogic as well as historical point of view.

3.5 Semantics

Besides the use of operators to deal with syntactical relationships, PRECIS includes a supplementary system for generating, from a machine-held thesaurus, references between semantically related terms, which might also function as a user's access points to the alphabetical index. It could be said that these references add a second dimension to the subject index. This separation of syntax from semantics in no way suggests that they are not interrelated. On the contrary, there exists a strong relationship between syntax and semantics, and on some occasions the indexer has to make a choice between the alternative ways in which certain relationships between concepts can be expressed in PRECIS (cf. Section 3.3.2.4). In fact, PRECIS possesses one special advantage so far as thesaurus construction is concerned. Experience has shown that it is a simpler matter to organize terms into logical classes within a thesaurus if the indexer works within the constraints of a grammar which defines exactly what is meant by a 'term'. We have already seen that the role operators play some part in this process, since they predicate that certain kinds of concepts should be treated as separate syntactical units, each prefixed by its own operator, when a string is written by the indexer. The 'Rules of differencing' can be specifically pointed out, which makes a major contribution to this aspect.
The PRECIS thesaurus is constructed in an inductive way, which means the process moves 'upwards' step-by-step from the individual term to the names of its genera and other related terms. Other thesauri are usually made deductively. The three different factors which form the basic components of the semantic side of PRECIS are:

(a) Indexing terms -- When building the thesaurus, only lead terms are taken into consideration. However, any new term, as soon as it appears into the lead position of an entry, is admitted into the network; that is to say, the vocabulary is open-ended.

(b) Reference Indicator Number (RIN) -- Each of such terms is assigned to an address in a random-access file in the computer, and is identified by seven-digit number (called Reference Indicator Number or RIN) which specifies this address.

(c) Relational codes -- The various kinds of relationship between terms held at different addresses (RINs) are indicated by a set of codes. These codes form part of the data associated with each term, and they generally express the following relationships, which are recognized in the ISO 'Guidelines for monolingual thesauri' (International Organization for Standardization, 1986):

i) *Equivalence relationship* (code $m$):

   Synonyms, e.g., Assessment/Evaluation
   Quasi-synonyms, e.g., Literacy/Illiteracy

ii) *Hierarchical relationship* (code $o$):

   Genetic relations, e.g., Plants
   Trees
   Whole-part relations, e.g., Great Britain
   England
   Systems and organs of body, e.g., Circulatory system
   Cardiovascular system
Areas of discourse, e.g., Behavioural sciences

Psychology

iii) Associative relationship (code $n, $x, $y):

Same category, e.g., Ships/Boats

Different category, e.g., Data processing/Computer systems

Once a first level of related terms has been established in the way quoted above, the indexer next considers each of the broader and associated terms. Each of these is considered as a potential indexing term, and the indexer goes on to establish a second level of semantically related terms. They too are assigned their respective RINs and relational codes. This operation continues till the indexer is satisfied with the hierarchy and has assigned it to the store. Once such a hierarchy has been constructed, the numbered address of any term in it can be quoted as part of the indexing data whenever the term occurs in an input string. The presence of RIN directs the computer to the appropriate address, and starts to generate See and See also references from all semantically related terms to the lead term in question. For example, the following set of cross-reference entries would be produced automatically by the computer from the term 'Assessment':

Evaluation See Assessment

Testing See Assessment

Measurement

See also

Assessment

Assessment

See also

Marking

Notes

1 However, it should be noted that, no attempt has been made to offer an
explanation in languages other than English (except occasional references), as it would be beyond the scope of this study as well as the comprehension of this author.

2 This view is contested by Michell (1979, p. 101), who claims that speech situations could be envisaged where the order of elements exemplified in an inverted word-order subject heading would be perfectly natural, and that such structure, while unusual in isolation, is certainly part of natural language structure. Inverted order occurs naturally, for example, in "There were books, illustrated and decorated, displayed all over the floor". Such examples demonstrate that indexing languages blamed for using unnatural word-order, in fact, use natural language structure as well, even if the structures they use are not those most frequently used or most simply described.

3 Hancox (1983) has written a translingual switching program which can translate a PRECIS string written in one NL (English) into its conceptual equivalent in a second (target) language (French), ready for manipulation into entries in the target language. The program incorporates a special routine for inserting lead-only terms into French strings in response to certain adjectival constructions in English, via a lexicon.


APPENDIX 3.1: PRECIS OPERATORS AND CODES*

SCHEMA OF OPERATORS

<table>
<thead>
<tr>
<th>Primary operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment of core operators</td>
</tr>
<tr>
<td>Core concepts</td>
</tr>
<tr>
<td>2 Action; Effect of action</td>
</tr>
<tr>
<td>3 Agent/Instrument/Intake/Factor</td>
</tr>
<tr>
<td>Extra-core concepts</td>
</tr>
<tr>
<td>5 Selected instance</td>
</tr>
<tr>
<td>6 Form of document; Target user</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinate concepts</td>
</tr>
<tr>
<td>g Standard coordinate concept</td>
</tr>
<tr>
<td>Dependent elements</td>
</tr>
<tr>
<td>q Member of a quasi-generic group</td>
</tr>
<tr>
<td>r Assembly</td>
</tr>
<tr>
<td>Special classes of action</td>
</tr>
<tr>
<td>t Author-attributed association</td>
</tr>
<tr>
<td>u Two-way interaction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme interlinks</td>
</tr>
<tr>
<td>$y$ 2nd/subsequent concept in theme</td>
</tr>
<tr>
<td>$z$ Common concept</td>
</tr>
</tbody>
</table>

* The schema of operators also include "Term codes" and "Typographic codes", which are not listed here.
<table>
<thead>
<tr>
<th>Secondary codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differences</td>
</tr>
<tr>
<td>Preceding differences</td>
</tr>
<tr>
<td>(3 characters)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

3rd character = number in the range 1 to 9 indicating level of difference

<table>
<thead>
<tr>
<th>Date as difference</th>
<th>$d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parenthetical differences</td>
<td></td>
</tr>
<tr>
<td>$n Non-lead parenthetical difference</td>
<td></td>
</tr>
<tr>
<td>$o Lead parenthetical difference</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Connectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v Downward-reading connective</td>
</tr>
<tr>
<td>$w Upward-reading connective</td>
</tr>
</tbody>
</table>
CHAPTER 4
INTRODUCTION TO DSIS

In sharp contrast to the kind of introduction to PRECIS provided in the last chapter, this chapter presents a rather formal description of the Deep Structure Indexing System (DSIS) or computerized POPSI [1]. The reason behind this decision is to give the reader a firsthand understanding of the system, which till this date has remained practically obscure in the Western hemisphere. The description is also devoid of any theoretical overtones, even though it supposedly has its origin in the classificatory principles of Ranganathan and his Indian school. Apart from Bhattacharyya's (1981b) brief description of the major contributions originating out of classificatory research in India, mainly during the post-Ranganathian era, a recent survey has tried to give an overview of the various facets of this group's work, especially in the field of verbal subject indexing (Biswas, 1988).

4.1 Subject Indexing Language

A Subject Indexing Language (SIL) is an artificial language used for formulating names of subjects. The term 'Subject Indexing Language' is now well-established as a scientific term, the denotation of which can be well-understood by constituent-analysis of the expression. For this purpose, each of the substantives, viz., 'Subject', 'Indexing' and 'Language', needs a definition of its own (Bhattacharyya, 1981a). A 'Subject' is essentially a piece of non-discursive information or a unit fact; and it is conveyed by an indicative formulation that summarises in its message what a particular body of information (document) is about (Bhattacharyya, 1981b; Devadason, 1986a). The process of preparing an index is 'Indexing', generating 'groups' of what they refer to. The purpose of 'Classification' as a process is also to
generate 'groups'. Teleologically, therefore, indexing is a process of classification. Bhattacharyya (1983b) has revealed that classification may be either i) 'Organizing classification', or ii) 'Associative Classification'. In organizing classification the classes are grouped on the basis of their 'COSSCO (Coordinate-Superordinate-Subordinate-Collateral) relationships', the result of which is always a hierarchy. In associative classification the hierarchy is not so explicit, since a group is formed due to the presence of a mutual common factor (say, alphabetical contiguity) among each of its members. The essential ingredients of a 'Language' -- natural or artificial -- are i) elementary constituents, and ii) rules for the formulation of admissible expressions (Bhattacharyya, 1979a). In this sense, a SIL is an artificial language with the following primary components:

i) the elements or vocabulary;
ii) the categorization of the elements on the basis of their semantic significance; and
iii) the rules of syntax with reference to the categories, for formulating admissible names of subjects, indicating the nature of relation between the elements (Bhattacharyya, 1983a; Devadason, 1986a).

4.2 Deep Structure of Subject Indexing Languages

For the purpose of designing a SIL the following types of structure in a name of subject have been recognised by Bhattacharyya (1979b, 1979c):

i) semantic structure, the structure in the dimension of denotation or comprehension, based on 'genus-species', 'whole-part', 'broader subject-narrower subject' relationships;

ii) elementary structure, the structure in the dimension of different 'categories' to which the different substantive constituents of subject-propositions belong; it is recognized on the basis of the
iii) syntactic structure, the structure originating from its rules of syntax which helps to preserve the meaning of the name of subject and to arrive at a consistent sequence of component elements.

The elementary structure and syntactic structure are closely related. Together they carry the responsibility of preserving the meaning of the name of subject. They are rather peculiar to the SIL concerned and are postulated for the purpose of information retrieval. On the other hand, the semantic structure is more concerned with the denotation of the index terms belonging to the particular natural language concerned. The structure of a specific SIL may be deemed to be a 'Surface Structure' of the 'Deep Structure of SILs'. By logically abstracting the structures of SILs of Cutter, Kaiser, Dewey and Ranganathan, a 'Deep Structure of SILs' has been arrived at by Bhattacharyya (1980, 1979c), which in essence parallels the idea of the 'Absolute Syntax' first postulated by Ranganathan (1967a), and further developed by Neelameghan (1971, 1979). The following diagram schematically represents the Deep Structure of SIL as postulated by Bhattacharyya:

![Deep Structure of SIL](image)

Figure 4.1: Deep Structure of SIL

[Annotation: D = Discipline; E = Entity; A = Action; P = Property; and m = Modifier, representing respectively the manifestations of the four Elementary Categories (EC) and the special component, Modifier. The
direction of the arrows indicate that any one of the ECs may be related to any one of another category provided the product of their relation is meaningful. Apart from these categories, there are two other concepts namely 'Base' and 'Core' which serve the purpose of bringing together all or a major portion of information pertaining to a manifestation or manifestations of a particular EC.

Besides the deep structure represented above, Bhattacharyya (1979c, 1982b, 1981b) also furnished a number of associated postulates as the essence of the general structural theory of linguistics of SIL. The basic sequence-based modulated chains complemented by cyclic permutation of sought components can endow a SIL with all these qualities. POPSI used this technique.

The Deep Structure Indexing System (DSIS), developed by Devadason, is based on the above 'Deep Structure' of SIL. It is based on: 1) a set of postulated Elementary Categories of the elements fit to form components of subject-propositions [2]; 2) a set of syntax rules with reference to the Elementary Categories; 3) a vocabulary control tool such as Classaurus; 4) a set of indicator digits to denote the Elementary Categories and their subdivisions; and 5) a set of codes to denote a few of the decisions of the indexer, in order to generate by computer manipulation, different types of subject indexes.

4.2.1 Elementary Categories

The Deep Structure of Subject Indexing Languages postulates that the component ideas in a subject-proposition can belong to any one of the Elementary Categories: Discipline, Entity, Property and Action, and a special component called Modifier.

**DISCIPLINE**: An Elementary Category that includes conventional fields of
study (branches of learning), or any aggregate of such fields or artificially created fields analogous to those mentioned above; e.g., Economics, Library and information science, Social sciences, Management science, etc.

ENTITY : An Elementary Category that includes manifestations having perceptual correlates, or only conceptual existence, as contrasted with their properties, and actions performed by them or on them; e.g., Personnel, Students, Libraries, Time, etc.

PROPERTY : An Elementary Category that includes manifestations denoting the concept of 'attribute' -- qualititative or quantitative; e.g., Skills, Productivity, Efficiency, etc.

ACTION : An Elementary Category that includes manifestations denoting the concept of 'doing'. Action may manifest as Self Action (refer to internal processes and intransitive actions) or External Action (capable of taking objects, i.e., transitive actions). For example, Migration, Hibernation, etc. are Self Actions; Evaluation, Indexing, Teaching, etc. are External Actions.

MODIFIER : In relation to a manifestation of any one of the Elementary Categories, 'Modifier' refers to an idea used or intended to be used to modify (qualify, differentiate) the manifestation without disturbing its conceptual wholeness, e.g., 'Adult' in 'Adult education', 'Skilled' in 'Skilled personnel', 'National' in 'National libraries', etc.

A Modifier (difference) generally creates a Species/Type of the modifyee (focus). Any manifestation of any Elementary Category may serve as the basis for deriving a Modifier. A Modifier can modify a single manifestation of any one of the Elementary Categories, as well as a combination of several
manifestations of more than one Elementary Categories. For example, in the subject-proposition 'Education in Great Britain', 'Great Britain' is a Modifier to the single manifestation of the Discipline 'Education'. On the other hand, in 'Education of women in Great Britain', 'Great Britain' is a Modifier to the combination of two manifestations: 'Education' (Discipline) and 'Women' (Entity). Modifiers can be divided into Common Modifiers and Special Modifiers. A Modifier capable of modifying manifestations of more than one Elementary Category, occurring singly or in combination, is a Common Modifier. DSIS admits of four different types of Common Modifier, viz., Form, Time, Environment and Place -- the examples of each being 'Bibliographies', '20th century', 'Deserts', and 'Great Britain', respectively. On the other hand, a Modifier having the potency of being used to modify manifestations of one and only one Elementary Category is a Special Modifier, which can be either Discipline based or Entity based or Property based or Action based. For example, 'Adult' in 'Adult education', 'Children' in 'Hospitals for children', 'Contagious' in 'Contagious diseases', 'Classification' in 'Grouping by classification', etc.

4.2.2 Subdivisions of Manifestation

Manifestations of each of the Elementary Categories may admit of subdivisions: Species/Type, Part and Constituent. A Species/Type does not disturb the conceptual wholeness of the manifestation to which it is a Species/Type. The relationship between a whole and its types is described as 'Genus-Species' relationship. A Part is a non-whole of the manifestation to which it is a Part. The relationship between a whole and its parts is described as 'Whole-Part' relationship. A Constituent is an ultimate part with its own individuality. For example, in the case of 'Computer', 'Micro-computer' is a Species/Type; 'Hardware', 'Software' are Parts; 'Silicon', 'Wire', 'Glass', etc are Constituents. Constituents generally occur for the
Elementary Category Entity.

4.2.2.1 Modifier: Compound and Complex Terms

Depending on the structure of the 'Modified Term', Modifiers could be further grouped into two types [4]:

1) Modifier of Kind 1, that which requires the insertion of a phrase or auxiliary words or function words between the modifyee term and its modifier term forming a Complex Term. For example, 'Library for blind' which is a type of 'Library'; and

2) Modifier of Kind 2, that which does not require a phrase or auxiliary words or function words to be inserted in between, but automatically forms an admissible Compound Term. For example, 'On-line information systems' which is type of 'Information systems'.

In DSIS Complex Terms formed using auxiliary/function words are also used to represent Complex Subjects. A subject formed by assembling two or more subjects expounding, or on the basis of, some relation between them constitutes a Complex Subject. Each component in such an assembly is called a 'phase' and the mutual relationship between the phases of an assembly is called 'phase relation'. Ranganathan (1967b, p. 358) introduced five kinds of phase relation -- General, Bias, Comparison, Difference and Application. In addition to the above five, Bhattacharyya (1979a, p. 18) prescribed two more, viz., Similarity and Application. Neelameghan and Gopinath (1969, 1972) have carried out detailed studies of phase relations. Complex Subjects formed by phase relations are generally narrower than the subject represented by the first phase. For example, 'Statistical methods biased to Librarians' is narrower than 'Statistical methods' in general and could be considered as a Species/Type. Moreover, several subjects considered as Complex Subjects formed by phase relation, could be later recognized to form 'Fused Subjects' represented by Compound Term/terms. For example,
'Information retrieval services using Computer systems', considered as a Complex Subject, could be recognized as a Fused Subject 'Computerized information retrieval services'. Due to the reasons mentioned above, phase relations are deemed to be formed by Modifiers of Kind 1, requiring insertion of auxiliary/function words denoting the types of relationship between the phases, forming Complex Terms.

The auxiliary/function words in Complex Terms may consist of role indicating words (as in 'Politics role of Women') or phase relation indicating words (as in 'Poverty influenced by Overpopulation') or prepositions (as in 'Management by Objectives) or other connectives. In DSIS, the incidence of these auxiliary/function words are limited only to the formation of Complex Terms. These function words also join in the arrangement of subject-propositions in the subject index formed according to this system. Thus, it is essential to standardize these auxiliary/function words to achieve consistency in the representation of subject-propositions.

4.2.2.2 Composite Term

According to the postulate of Elementary Categories (see Section 2.1), any one component idea in a subject-proposition can belong to any one and only one of the Elementary Categories. If a component term represents manifestations of more than one Elementary Category then it is a Composite (Category) Term. It should be factored into two or more constituent terms or elemental concepts (Soergel, 1974, p. 74) and each one of them should be identified as belonging to one or the other of the Elementary Categories. While factoring Composite Terms, one must be careful to ensure that the correct meaning of the Composite Term is provided by the combination of the factored component terms. Also care must be taken while making distinction between Compound Terms and Composite Terms. For example, 'Library
management' is a Composite Term which is to be broken down into 'Libraries (E) + Management (A)'. But 'Management libraries' is a Compound Term which is a Species/Type of 'Libraries' both being manifestations of the same single Elementary Category Entity. The identification and factoring of Composite Terms is to be decided in the context of the subject-proposition as a whole and the Elementary Categories of the indexing language. The Composite Term is considered in DSIS as a synonymous term to the combination of the factored constituent terms and a Cross Reference (CR) entry (see Section 4.3) is included in the subject index referring the user from the Composite Term to the combined factored terms, e.g.,

Pneumonia = Medicine (D) + Lung (E) + Bacterial infection (P).

4.2.3 Syntax of DSIS

The basic rule of syntax associated with the Deep Structure of Subject Indexing Languages for formulating subject-propositions is that, Discipline should be followed by Entity (both modified or unmodified) appropriately interpolated or extrapolated wherever needed by property and/or Action (both modified or unmodified). A manifestation of Property follows immediately the manifestation in relation to which it is a Property. A manifestation of Action follows immediately the manifestation in relation to which it is an Action. Property and Action can have another Property and/or Action directly related to them. A Species/Type or Part or Constituent, follows immediately the manifestation in relation to which it is a Species/Type or Part or Constituent. A Modifier follows immediately the manifestation in relation to which it is a Modifier. If there are more than one Modifier to the same manifestation, any one valid sequence of them in terms of their representation in the natural language concerned is acceptable. In other words, if the modified term forms a Complex Term, it should form an acceptable natural language title-like phrase, and if it forms a Compound
Term it should be an admissible one in the natural language concerned. As per the above rules of syntax, only the positions of Discipline and Entity are fixed. The positions of both Property and Action are not fixed, rather they are given a 'floating' position.

For example, consider the following name of subject, 'In economics, control of money circulation'. The Discipline is 'Economics'. The Entity is 'Money'. There are two Actions: 'Circulation' and 'Control'. 'Circulation' is an Action on the Entity 'Money' (it is 'Money' which is circulated). Hence, it should appear adjacently next to it. The Action 'Control' is not operative unless the concept of the Action 'Circulation' is conceded. It is the 'Circulation' process which is controlled, and it should go adjacently next to it. Hence, the formalised name of subject according to the rules of syntax would be:


In general, the above rules of syntax give rise to the following syntactical structure to a name of subject formulated according to the Deep Structure of Subject Indexing Languages:

"DISCIPLINE followed by ENTITY which is followed by PROPERTY and/or ACTION. PROPERTY and/or ACTION may be further followed by PROPERTY and/or ACTION as the case may be, followed by COMMON MODIFIERS. The SPECIES/TYPES and/or MODIFIERS and/or PARTS and/or CONSTITUENTS, for each of the ECs follow immediately adjacent to the manifestation to which they are respectively SPECIES/TYPES or MODIFIERS or PARTS or CONSTITUENTS without the manifestation of any other EC intervening" (Devadason 1986a, p. 5).

A schematic representation of the syntactic structure of the DS of SIL is shown in Figure 4.2 wherein, the directions shown by the arrows indicate that any manifestation of any of the Elementary Categories may be related to any of another category provided their representation is meaningful and is
according to the rules of syntax. These rules of syntax give rise to a context-dependent sequence of the components in the name of subject in

\[
\text{DISCIPLINE} \downarrow \text{ENTITY} \downarrow \text{PROPERTY (and/or)} \downarrow \text{ACTION} \downarrow \text{COMMON MODIFIER}
\]

\[
\text{-smp} \quad \text{-smpc} \quad \text{-smp} \quad \text{-smp} \quad \text{-smp}
\]

\[
\text{ACTION (and/or)} \downarrow \text{PROPERTY (and/or)} \downarrow \text{ACTION}
\]

\[
\text{-smp} \quad \text{-smp}
\]

[where \text{s} = \text{Species/Type}, \text{m} = \text{Modifier}, \text{p} = \text{Part}, \text{c} = \text{Constituent}]

Figure 4.2: Syntactic Structure of DS of SIL

conformity with Ranganathan's (1967b, pp. 425-429) Principles for Facet Sequence -- the Wall-Picture principle and its corollaries such as the Actand-Action-Actor-Tool principle. Besides the above mentioned criteria of natural language acceptability, it is also helpful to follow these Principles in deciding the sequence of Modifiers too. For example, in the Complex Term 'Satellites (for) remote sensing (using) infra-red cameras' the sequence of Modifiers of Kind 1, viz., 'Remote sensing' and 'Infra-red cameras' are according to the Wall-Picture Principle in the sense that, unless and otherwise the type of activity the Satellite is supposed to perform (Remote sensing) is determined, the instrument to be used (Infra-red cameras) for performing it cannot be considered. The following example gives a series of Compound Terms formed by successive Modifiers of Kind 2, their sequence being in conformity with the Wall-Picture Principle: 'Students, School students, Elementary school students'. In both the cases the terms are acceptable in the natural language concerned.

It has been observed that, such an alphabetically arranged group of
subject propositions, formulated as per the DS of SIL following the rules of syntax, would provide a kind of organizing sequence among its components. All the subject propositions belonging to the same Discipline would be together. Within them, all belonging to the same Entity would be together and so on. This basic sequence can be manipulated to generate other different organizing sequences. Indexing (classification) is always purpose oriented; and it is the specific purpose that determines the optimally efficient and effective version of the grouping. When the purpose is to group together manifestations of a particular EC or combinations of ECs, the EC or combination of ECs, as the case may be, can be taken to be the first context specifying category instead of Discipline and called the Base. When the purpose is to further group together manifestations of any one EC, the Category concerned can be taken to be the second category following the Base, called the Core. This rule allows an indexer to decide which of the ECs or combination of them should become the first context specifying category and which EC is the second context specifying category, respectively. Once this is decided, the other rules of syntax guide the formulation of subject propositions to produce the grouping required.

4.2.4 Indicators of Deep Structure

In order to reflect the DS of SIL in the formulated subject propositions, the following numeric codes have been used in DSIS to indicate the manifestations of the different ECs, their subdivisions and Modifiers of different kinds:

<table>
<thead>
<tr>
<th>Common Modifiers</th>
<th>Elementary Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Form Modifier</td>
<td>9 Discipline</td>
</tr>
<tr>
<td>2 Time Modifier</td>
<td>8 Entity</td>
</tr>
<tr>
<td>3 Environment Modifier</td>
<td>.2 Property</td>
</tr>
<tr>
<td>4 Place Modifier</td>
<td>.1 Action</td>
</tr>
</tbody>
</table>
Subdivisions/Divisors

.3 Constituent
.4 Part
.5 Modifier of Kind 1 including Phase Relation Modifier
.6 Species/Type, including those created by Modifier of Kind 2.

In a subject statement the indicators precede the components to which they are indicators. The indicators for Property and Action, and also for the Subdivisions/Divisors are attached with the indicators for the ECs to which they are respectively Property or Action or Subdivisions/Divisors. The component terms along with their respective indicators formed in the sequence as per the rules of syntax, constitutes the name of subject formulated as per DSIS, which is illustrated in the following section.

4.3 Formulation of Name of Subject

4.3.1 Content Analysis and Formalisation

The formulation of the name of a subject in DSIS starts with writing out sentences such as "this document (book, report, article, etc.) is about...". To aid in writing out such an indicative formulation that summarises in its message 'what a particular body of information is about', the title of the document is taken as the starting point, which is further enriched by additional information from other parts of the document such as the contents, index, abstract, etc., if necessary, the text also. For example, we have a subject statement "Use of personal computers for searching legal databases". Each of the component ideas corresponding to the ECs such as the name of the Discipline, the core Entity of study, etc., that are implied in the expressed subject statement are explicitly stated to form an 'expressive title-like phrase' as follows: "In information retrieval, use of personal computers for searching legal databases". The expressive title-like phrase is then analysed to identify the components of the DS of SIL -- the ECs,
their Species/Types, Parts, Constituents and Modifiers of different kinds -- to which each of the components in the title-like phrase belong. Composite Terms, if any, which are not amenable to such categorization are broken down into their elemental concepts and then identified accordingly. Such Composite Terms and their corresponding elemental concepts are to be noted separately to formulate Cross Reference entries for inclusion in the index during final sorting and printing. A standard indexing and vocabulary control tool such as thesaurus or Classaurus is to be used for this purpose. It is possible to construct the Classaurus beforehand, i.e., before applying it to indexing, or it can be constructed along with indexing work (Bhattacharyya, 1981b, p. 17). Following the latter procedure, an attempt has been made in this project to construct an online alphabetic Classaurus using the DSIS index strings as input. Following the rules of syntax the component terms are written down as a formalised expression in a context-dependent sequence. The analysed and formalised statement of our subject proposition is given below:


4.3.2 Modulation and Standardization

Each of the component terms in the analysed and formalised subject proposition is then 'modulated' by enhancing it with successive interpolating and/or extrapolating superordinates of each EC manifestation, by finding out 'of which it is a Species/Type or Part or Constituent'. The superordinate terms recognized are prefixed to (extrapolated to the left of) the term under consideration. This process of finding out the superordinates is continued with each of such superordinates recognized in the process till it ends up with the concept of the EC of which it is a manifestation.
In our example above, the first term represents the name of the Discipline. Since, it has been taken as the starting point, there is no need to 'modulate' it. The next term 'Legal databases' represents the EC Entity. In the context of 'Information retrieval' it is a type of 'Databases', which is again a type of 'Information sources', both being superordinate terms to their subordinate terms and fixed prior to the latter. The term 'Information sources' does not have any superordinates in the context of the subject under consideration and ends up with the concept Entity. The next component term is 'Searching' which does not have any superordinate term as such. But it is being modified by the Entity term 'Personal computers', which is a type of 'Computers', a superordinate term to the former [5]. Hence, the modulated subject proposition would be as follows:

"(D) Information retrieval, (E) Information sources, (Type of E) Databases, (Type of E) Legal databases, (A on E) Searching, (Type of A) (use of) Computers, (Type of A) Personal computers".

For the purpose of modulation, vocabulary control tools such as classauri, thesauri, and other terminological sources such as glossaries, dictionaries, including classification schemes can be used. Besides, helping to represent exactly the depth or intension of the subject, the other reasons for fixing the successive superordinates prior to each of the respective manifestations is to establish their proper context, make their denotation in the subject statement precise and unambiguous, to endow the subject statement with the capacity to produce an organizing sequence effect resembling the sequence of the expressive notation of a class number. Moreover, it is possible to prepare an alphabetical subject index using all or the relevant superordinate terms as Lead Terms, from the modulated name of subject itself.

After the 'modulation' step, each of the component terms in the name of
subject, including the auxiliary/function words introducing Modifiers of Kind 1, are substituted by standard terms and the synonymous and quasi-synonymous terms are noted separately for preparing Cross Reference entries to be included in the index later. A standard thesaurus or Classaurus is to be used to accomplish this process of 'standardization'. In our example, we can replace the term 'Personal computers' with the standard term 'Microcomputers'. It is also necessary to standardize the auxiliary/function words introducing Modifiers of Kind 1, as these participate in the arrangement of names of subjects. Thus we get the standardized name of subject as follows:

"(D) Information retrieval, (E) Information sources, (Type of E) Databases, (Type of E) Legal databases, (A on E) Searching, (Type of A) (using) Computers, (Type of A) Microcomputers".

4.3.3 Category Indication

Appropriate indicators for ECs, their subdivisions/divisors and Common Modifiers of different kinds are inserted in the appropriate places. The auxiliary/function words denoting the Modifier of Kind 1, if any, enclosed within parenthesis, are prefixed to the Modifiers of Kind 1 wherever warranted. All other auxiliary words denoting the different ECs, their subdivisions/divisors and Common Modifiers are removed. The resulting name of subject is as follows:

"Information retrieval 8 Information sources 8.6 Databases 8.6 Legal databases 8.1 Searching 8.1.5 (using) Computers 8.1.6 Microcomputers".

In this subject proposition, the indicator for Discipline is not used as it is taken as understood to be the first digit in all subject statements. In the component '8 Information sources', the indicator '8' denotes that it is a manifestation of Entity. In the component '8.6 Databases', the indicator '8.6' denotes that it is a Species/Type of Entity 'Information sources'. In
the component '8.6 Legal databases', the indicator '8.6' denotes that it is a Species/Type of the Type of Entity 'Databases'. The component term 'Legal' is a Modifier of Kind 2 creating the Species/Type 'Legal databases'. In the component term '8.1 Searching', the indicator '8.1' denotes that it is an Action on Entity. Similarly, in '8.1.5 (using) Computers', the indicator '8.1.5' denotes that it is a Modifier of Kind 1 to the Action on Entity and so on.

A set of 'modulated and standardized' subject propositions with appropriate indicators when just alphanumerically can produce an organizing classification effect as mentioned earlier. This organizing effect has reduced considerably the See also Cross References referring from narrower subjects/terms to their respective broader subjects/terms (ascending references) and from broader subjects/terms to their respective narrower subjects/terms (descending references).

4.4 Subject Index Entry

An 'index entry' refers to an indexed item somewhere outside the index, say, to a book on the library shelf, to a record in a database (Craven, 1986, p. 1). An index entry may be divided into a 'description' consisting of index term (s) and a 'locator' or 'address'. An index entry specifying a subject along with its address is a 'subject index entry'.

4.4.1 Functions of Subject Index Entry

Subject index entries in general, perform three functions (Keen, 1977, p. 19):

1) locating function -- to permit the location of entries for the subject;

2) comprehending function -- to give data for the comprehension of the
entries to permit relevance prediction; and

3) organizing or relating function -- to aid the location of entries for subjects related to the one being sought.

4.4.2 Structure of Subject Index Entry

In DSIS, a subject index entry consists of a 'Lead heading' at its beginning with a 'Lead Term' occupying the first position in it. The Lead Term mainly decides the location of an entry in the subject index, thus catering to its locating function mentioned above. The other constituents of a Lead Heading may consist of other index terms specifying (qualifying) the Lead Term. Such terms are formed by using terms that are Upper Links (superordinates belonging to other ECs as well as occurring earlier in the subject proposition) to the Lead Term concerned, and are called Upper Link Specifiers (known as 'Qualifier' in PRECIS). Lead Headings do not contain any indicators, as these generally cater only to the locating function.

A subject index entry may contain a 'Context Heading' (same as 'Display in PRECIS) occupying the immediate next section to the Lead Heading section. It consists of index terms providing context to the Lead Heading and aids the user to judge the relevance of the entry through comprehension of its meaning. The Context Headings usually express the complete subject analysis of the document being indexed. These form sub-entries to the same Lead Heading and are used for systematic grouping by bringing together related subjects having the same Lead heading. In this sense, Context Headings perform the organizing and relating function of a subject index entry. For this purpose of producing an 'organizing classification effect' the Context Headings in DSIS also contain the indicator digits in them.

Subject index entries (other than Cross Reference entries) generally contain the address or location of the document being indexed for its unique
identification. It may be an index number or record number. Instead of the
index number or record number, the full bibliographic details of the source
of information may also be given.

Thus, a subject index entry in DSIS consists of three distinct
sections, viz., the Lead Heading Section, the Context Heading Section, and
the Location or Address Section. These are shown in Figure 4.3 below.

<table>
<thead>
<tr>
<th>Lead Heading Section</th>
<th>Lead Term, Upper Link Specifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context Heading Section</td>
<td>Context Terms According to Syntax</td>
</tr>
<tr>
<td>Location/Address Section</td>
<td>Index No.</td>
</tr>
</tbody>
</table>

Figure 4.3: Structure of Subject Index Entry

4.4.3 Cross Reference Entry

To bring together subjects related to the Lead Heading, Cross Reference
entries directing the user from one Lead Heading to another are used in most
indexing systems. Such Cross Reference entries also cater to the relating
function of a subject index entry. These are of two types -- See and See
also -- directing the user from the 'referred from' heading to the 'referred
to' heading. In DSIS, Cross Reference entries using the 'directing element'
See are used to refer the user from a non-standard (synonymous or quasi-
synonymous or variant form of the) heading to the standard heading. For
example,

- CIP See Cataloguing (in) Publication
- Accuracy See Precision
- Encyclopedia See Encyclopaedia

In DSIS, there are no See also Cross Reference entries of the 'descending'
(broader to narrower subjects) or 'ascending' (narrower to broader subjects)
types. Neither are there any of the 'associatively' related type, because,
what is associatively related to what, and how they are related is revealed by the Context Headings. However, *See also* Cross Reference entries may be needed in very special circumstances.

In DSIS, apart from Cross Reference entries used to control naturally occurring synonyms, quasi-synonyms, etc., *See* Cross Reference entries are needed to control synonyms artificially created by factoring Composite Terms. For example,

- Library management *See* Libraries (E) + Management (A)
- Pneumonia *See* Medicine (D) + Lung (E) + Bacterial infection (P)

All the synonyms, quasi-synonyms, synonyms due to factoring etc., can be ascertained by referring to the alphabetical and systematic parts of the Classaurus for the concerned Discipline/Base.

Cross Reference entries may be presented as a single-line uni-section entry (as in the examples given above), or as a tri-section entry (as given below):

- Digital computer systems
  - *See*
  - Computer systems
- Price elasticity
  - *See*
  - Price (P) + Elasticity (P)

In DSIS, as it is not necessary to provide *See also* Cross Reference entries of the ascending or the descending types, their number would be less and it may suffice to have the uni-section entry only.

4.4.4 Permuted Cross Reference Entry

In DSIS, a special type of Cross Reference entries called 'Permuted Cross Reference entries' is formed by cyclic permutation of constituents in a
Complex Term (formed by Modifier of Kind 1), so that significant constituent terms in it also form the Lead. These entries are constituted only of Lead Headings (i.e., no Context Heading section and Location section) having a changed sequence of the constituent terms to the standard rendering of the Complex Term as Lead terms. These permuted entries arising out of a Complex Term do not contain any other section such as the 'directing' section or 'referred to' section. Instead, they contain a '/' (backward slash) to indicate the beginning of the Complex Term so that the entry could be read from that position onwards in a cyclic wrap around manner to its end, which is again indicated by the same '/', giving the standard rendering of the Complex Term. This feature produces entries somewhat like KWIC (Keyword-in-context) entries. Other sections such as Context Heading section and Location section are provided only under the standard rendering of the Complex Term. Due to this function of referring from a variant, permuted form of a Complex Term to its standard rendering, these entries are called Permuted Cross Reference entries. All the significant constituent terms in a Complex Term including the Modifier of Kind 1 and even a constituent of such a Modifier are selected to form the Lead in Permuted Cross Reference entries. However, the auxiliary/function words in a Complex Term are not led. Consider the following name of subject formulated according to DS of SIL:

Information retrieval 8 Libraries 8.6 Academic libraries 8.4 Patrons 8.2 Skills 8.2.6 Search skills 8.2.5 (influenced by) Online bibliographic databases

The Complex Term in the above name of subject is:

Search skills 8.2.5 (influenced by) Online bibliographic databases

In this case it is necessary to have Leads under both the Modifier of Kind 1 'Online bibliographic databases' and its constituents 'Bibliographic databases' and 'Databases'. The Permuted Cross Reference entries that will
be generated from the above string are as follows [6]:

Online bibliographic databases / Search skills (influenced by)
Bibliographic databases / Search skills (influenced by) Online
Databases / Search skills (influenced by) Online bibliographic
In special situations the conjunction 'and' may also be used in a Complex Term to form a multi-focal component, e.g.:

Distance learning 8.1.5 (using) Televisions 8.1.5 (and) Videos
Both the constituent terms 'Televisions' and 'Videos' can be selected to form the Lead in Permuted Cross Reference entries generated from the above Complex Term, to cater for searches using the terms 'Televisions' or 'Videos'. The entries that will be generated are:

Televisions (and) Videos / Distance learning (using)
Videos / Distance learning (using) Televisions (and)
In DSIS, multi-focal or multi-theme documents are treated as separate names of subject for each theme. The only exception to this rule is the case mentioned above.

4.5 Formation of Subject Headings

After formulating the name of subject as per DS of SIL, standardizing the terms in it and noting down the necessary Cross Reference entries, the indexer has to decide

1 which component terms should form the Lead; and
2 which component terms should form the Context,
in order to produce headings for the different subject index entries.

4.5.1 Selection of Lead Term

In order to provide access and to cater to the locating function of the index entries in DSIS (as in all indexing systems), significant or sought terms in the name of a subject are usually selected to form Lead. Several
Lead Headings are prepared from a particular name of subject, providing access through each of the Lead Terms. Despite the fact that it is very difficult to ascertain which terms are significant and which are not, general decisions concerning Lead Terms are to be formulated within an organisation and recorded as policy statements. For example, the term denoting the Discipline/Base need not be selected to form Lead, if the whole subject index is specifically for that Discipline/Base alone. Also, very generic Entity terms such as, man, animals, plants, etc., very common Action terms such as, evaluation, analysis, determination, etc., very common Property terms such as, efficiency, property, effectiveness, etc., and terms denoting Common Modifiers, viz., Form, Place, Time, etc., need not be necessarily selected to form Lead Terms. But if such common Entity, Property and Action terms have been modified and have given rise to Species/Type, then it may be necessary to select the terms denoting the Species/Type as Lead Terms. For instance, if the database to be indexed is exclusively on 'Consumer economics', then there is no need to give 'Consumer economics' as Lead term in a subject index entry to it; similarly, common Entity term like 'Commodities' may not form Lead, except when it is modified by such terms as 'Luxury', 'Essential' etc., and forming Species/Type such as 'Luxury commodities', 'Essential commodities', etc. Policies of this kind may vary from one organisation to another depending on the purpose of the subject index -- the subject area concerned, the material being indexed and the community of users to be served by the index. It is a precondition in DSIS that, each of the Complex Terms in the name of subject in their standard rendering form must be selected as Lead Terms to generate meaningful entries under subsequent terms following them. This is due to the reason that Permuted Cross Reference entries arising out of Complex Terms do not contain any Context Heading section and Location section as such and their only function is to refer back to the standard rendering of the Complex Term (see
The component terms selected as Lead Terms are prefixed with a 'process code $0'$ for computer manipulation (see Section 5.5).

### 4.5.2 Selection of Context Term

In order to help the user in comprehending the meaning of the subject denoted and to predict its relevance, Context Headings are provided in DSIS index entries. The Context Heading sets the context in which the Lead Heading occurs. To make the entries coextensive with the thought content of the document being indexed, Context Headings in DSIS represent the complete subject analysis. Moreover, in DSIS, the Context Heading also performs the 'relating function' by providing an organising sequence among the Context Headings to a particular Lead Heading when just sorted alphanumerically. In order to create the maximum possible organising effect (or collocation), all the component terms in the subject-proposition including those interpolated/extrapolated at the 'modulation' step, formalised and arranged according to the rules of syntax, along with the indicators of the respective Elementary Categories, their Species/Types, Parts, Constituents and Modifiers of different kinds should be kept in the Context Heading. For example,

**In-service training, Unskilled personnel, Industries, Labour economics**

Labour economics 8 Industries 8.4 Personnel 8.6 Unskilled personnel 8.1 Training 8.1.6 In-service training 0 Bibliographies

Labour economics 8 Industries 8.4 Personnel 8.6 Unskilled personnel 8.1 Training 8.1.6 In-service training 0 Inquiry reports 4 Africa

Labour economics 8 Industries 8.4 Personnel 8.6 Unskilled personnel 8.1 Training 8.1.6 In-service training 8.1.5 (by) Managers

In order to avoid any possible ambiguities arising out of variations in the sequence of terms in other parts of the index entry (such as the Lead Heading), the sequence of component terms in the Context Heading is kept invariant in DSIS, along with the different indicators. For example,
If the purpose of providing the Context Heading in a subject index entry is only to help the user in comprehending the meaning of it, then the superordinate (broader) terms interpolated or extrapolated at the 'modulation' step may be omitted, provided the full meaning is represented in it. Accordingly, the last component term (it may be a Compound Term or a Complex Term) in each of the manifestations of the Elementary Categories and Common Modifiers is selected to form such a 'Short Context Heading'. For example, the name of subject given above would give rise to the following 'Short Context Heading':

Labour economics 8 Industries 8.4 Unskilled personnel 8.1 In-service training 0 Bibliographies

But, if the chosen last component term of an Elementary Category does not by itself individualise it, then successive broader terms should also be selected so that it gets individualised and homonym free. In most of the cases, the term denoting the Discipline/Base is also selected to form the first context specifying category in the name of subject, because it has the highest potency in resolving homonyms.

It is also recommended that, while selecting a term denoting a Part or Constituent to form Context, it is worthwhile to select the immediate broader term representing its 'whole' also to form Context. This is because
the same Parts/Constituents may occur as such in different 'wholes' and would need individualisation by their respective wholes. Homonyms created by polyhierarchies, i.e., a single term belonging to two different hierarchies, should be resolved by incorporating the successive superordinate terms to it in the Context. For example, in 'Labour economics', 'Unskilled personnel' is a Part of 'Industries' and also of 'Agriculture', 'Commerce', etc. In order to make the Context Heading represent the meaning clearly, it is essential to select the terms denoting 'whole' Entity occurring in the concerned name of subject also to form the Context in such cases, such as:

Labour economics 8 Agriculture 8.4 Unskilled personnel 8.2 Wages
Labour economics 8 Commerce 8.4 Unskilled personnel 8.1 Training
Labour economics 8 Textile industries 8.4 Unskilled personnel 8.2 Skills

4.5.3 Upper Link Specifiers to Lead Term

A name of a subject formulated according to the DS of SIL can be considered as a Chain [6] having as links each of the component terms (Compound Terms and Complex Terms each taken as a unit component term). For a particular component term, all the other terms occurring earlier to it, may form Upper Links. In DSIS, Upper Links of not only the EC of the component term under consideration but also component terms belonging to other ECs arranged according to the rules of syntax, are considered as well. For instance, in the following Chain, for the term 'In-service training', each one of the terms occurring above it are Upper Links.

```
Labour economics (D)
| Industries (E)
| Unskilled personnel (Part of E)
| In-service training (A)
```

Figure 4.4: Chain and Links in a Name of Subject
When a term becomes the Lead Term, some of the Upper Links are suffixed to it to further specify the Lead Term as well as provide some sort of context to it. The Upper Links so used are called 'Upper Link Specifiers', similar in nature to 'Qualifier' in PRECIS. While selecting terms for forming Context Heading, care is taken to see that the terms selected are such that the Short Context Heading is unambiguous and homonym free (See sec. 5.2). Hence, the Upper Links to a term under consideration, that are selected to form Short Context Headings, are used to form Upper Link Specifiers to the concerned term when it becomes the Lead, e.g.

Students, Education

Education 8 Students 8.6 Handicapped students 8.2 Behaviour

8.2 Learning behaviour 8.2.1 Assessment

Handicapped students, Education

Education 8 Students 8.6 Handicapped students 8.2 Behaviour

8.2 Learning behaviour 8.2.1 Assessment

Learning behaviour, Handicapped students, Education

Education 8 Students 8.6 Handicapped students 8.2 Behaviour

8.2 Learning behaviour 8.2.1 Assessment

Assessment, Learning behaviour, Handicapped Students, Education

Education 8 Students 8.6 Handicapped students 8.2 Behaviour

8.2 Learning behaviour 8.2.1 Assessment

Apart from specifying the Lead Term, the inclusion of Upper Links in the Lead Heading helps the user to minimize his search effort. When the database being indexed is large, there would be several Context Headings under a single Lead Term. In order to judge the relevance, the user would have to scan through a large number of Context Headings to the particular Lead Term sought by him. If, along with the Lead Term, some of its Upper Links are kept to its right hand side forming a Lead heading, then the number of Context Headings having the same combination of Upper Links would
be small. For example, a comparison of the following two sets of entries will prove our point:

**Set 1 (Lead Term only):**

Adults
- Education 8 Adults 8.2 Abilities 8.2.6 Numerical abilities
- Education 8 Adults 8.2 Behaviour 8.2.6 Reading behaviour
- Psychology 8 Adults 8.6 Handicapped adults 8.2 Attitudes
  - 8.2.5 (to) Women
- Sociology 8 Adults 8.2 Alcoholism 8.2.1 Prevention

**Set 2 (Lead Terms with Upper Link Specifiers):**

Adults, Education
- Education 8 Adults 8.2 Abilities 8.2.6 Numerical abilities
- Education 8 Adults 8.2 Behaviour 8.2.6 Reading behaviour

Adults, Psychology
- Psychology 8 Adults 8.6 Handicapped adults 8.2 Attitudes
  - 8.2.5 (to) Women

Adults, Sociology
- Sociology 8 Adults 8.2 Alcoholism 8.2.1 Prevention

The sequence of component terms in the Lead Heading containing Upper Link Specifiers taken from left to right, is the reverse of the sequence of the component terms in the name of subject formed as per the rules of syntax of DS of SIL and is called 'Reverse Rendering' (Ranganathan, 1964b, p. 114). Conversely, 'Forward Rendering' is the sequence in which the component terms are arranged according to the rules of syntax of DS of SIL which is used in the Context Headings.

4.5.4 Default Lead and Context

Certain default Lead and Context selection options are available in DSIS to make it easier for the indexer and economize the cost of input. However,
these could be varied depending on the need of the individual indexing agency in question and the index generation program used for this purpose. Therefore, at this point, these may not be worth discussing, as they do not contribute to the basic understanding of the system.

4.5.5 Processing Codes

A set of 'processing codes' are used by the indexer in DSIS for computer manipulation of the name of subject to produce different Lead and Context Headings. These are used along with the indicators to denote different ECs, their Species/Types, Parts, Constituents and Modifiers of different kinds:

1) '$0' -- Lead Term;
2) '$1' -- Context Term;
3) '<' (starter), '>' (arrester) -- enclosed within, is a Complex Term;
4) '$2' -- Lead in PCR entries arising out of Complex Term;
5) '$*$' (auxiliary word identifier), '/' (auxiliary word delimiter) -- enclosed within, is an auxiliary/function word (s); and
6) '$9' -- neither Lead nor Context.
7) '$3' -- used with Modifiers of Kind 2 only, for automatic creation of Compound Terms.

4.6 Coding of Names of Subjects

The name of subject formulated as per the DS of SIL is augmented with the process codes described above to formulate the input string. Certain coding conventions are followed while assigning the process codes to individual component terms.

4.6.1 Coding Conventions

In writing an input string the following sequence should be followed:
Indicator, Process code, Subject component.

The process codes 'SO' and '$1' cannot be coded within a Complex Term enclosed within angular brackets, i.e., they have to be coded preceding the starter angular bracket of the Complex Term concerned. Similarly, the process code '$2' indicating Lead Terms in PCR entries, cannot be coded outside Complex Terms enclosed within angular brackets. The process code '$3' cannot be prefixed to the first constituent of a Complex Term given within angular brackets. Any number of Complex Terms can occur in a name of subject, but each should be separately enclosed within matching angular brackets. Every name of subject should begin with a '$' sign. To indicate the first component term in the name of subject being neither Lead nor Context, it should be prefixed with the null process code '$9'. Word forms of dates and other numerical values rather than numbers should be preferred in the name of subjects. However, if necessary, numerals may be included in the input string, but they must be prefixed with '$*$' and delimited by '/' after their applicable EC indicators. Processing codes should be written in the following sequence:

$0, $1, $2, $3; $9, $*

In the event of a clash, the latter occurring process codes take precedence over the former. The default punctuation mark separating component terms in the Lead Heading is a comma. There are no special punctuation marks in the Context Heading.

4.6.2 Coding Illustration

Consider the category-indicated, formalised name of subject (earlier being 'modulated' and 'standardised') given in section 3.3 as our example:

"Information retrieval 8 Information sources 8.6 Databases 8.6 Legal databases 8.1 Searching 8.1.5 (using) Computers 8.1.6 Microcomputers".

In order to form the Context Heading (Short Context Heading), it is
sufficient if the terms 'Legal databases', 'Searching (using) computers' and 'Microcomputers' are selected. As the term 'Information sources' itself resolves any homonym that may arise, it is not necessary to select the Discipline term 'Information retrieval' to form the Short Context. These terms are prefixed with the process code '$1', indicating that they form (Short) Context Heading. If we assume that the subject index is only for 'Information retrieval', then it is not necessary to select it to form the Lead either. Hence, it is prefixed with the null process code '$9'. To form Lead Terms, 'Information sources', 'Databases', 'Microcomputers' and the Complex Term 'Searching (using) Computers' are selected. These terms are prefixed with the process code '$0' to indicate that they are Lead Terms. The Complex Term is enclosed within angular brackets and the auxiliary/function word '(using)' between '$*$' and '/'. To form Lead Terms in PCR entries arising out of the Complex Term, the constituent term 'Computers' is selected and it is prefixed with the process code '$2'. Incorporating these decisions of the indexer, we get the following coded input name of subject:

"$9 Information retrieval 8 $0 Information sources 8.6 $0 Databases 8.6 $0$1 Legal databases 8.1 $0$1 <Searching 8.1.5 $* (using)/ $2 Computers> 8.1.6 $0$1 Microcomputers".

The above input string could be further simplified and economized by taking default options wherever applicable (see section 5.4) and using the process code '$3' for Modifier of Kind 2 to form Compound Term, as shown below:

"$9 Information retrieval 8 $0 Information sources 8.6 $0 Databases 8.6 $3 Legal 8.1 <Searching 8.1.5 $* (using)/ $2 Computers> 8.1.6 Microcomputers".

Besides these, the synonyms, quasi-synonyms and synonyms due to 'factoring' of Composite Terms are to be noted separately to form CR entries to be included in the index before final sorting and printing. For example,

Home computers = Microcomputers
4.7 Types of Indexes

The coded input name of subject is manipulated by computer to generate different types of Lead Headings and Context Headings [7]. In the formation of Lead Headings, two major types are possible. They are:

1. Lead Headings containing as Lead Terms only those component terms marked to form Lead; and
2. Lead Headings containing all the component terms (excluding constituent terms of Complex Terms and auxiliary/function words) as Lead, irrespective of the indication for Lead selection.

Within each of the above two major types, two further types of Lead Headings could be formed. They are:

1. Lead Headings containing only uni-component terms (i.e., terms between two consecutive EC indicators or between two consecutive starter and arrester angular brackets) as Lead Terms; and
2. Lead Headings containing Lead Terms and their applicable Upper Link Specifiers.

Similarly, in the formation of Context Headings, two different types are possible. They are:

1. 'Full Context Heading' containing all the component terms in the input name of subject along with their EC indicators, irrespective of whether they are selected to form Context or not; and
2. 'Short Context Heading' containing only the component terms selected to form Context along with their appropriate EC indicators.

By varying the combinations between the above different Lead and Context types, altogether fourteen different types of subject index entries could be automatically generated from a single coded input name of subject in DSIS, out of which any one could be selected as appropriate, depending on the
purpose at hand. These include POPSII entries, Chain Index entries and a
pseudo PRECIS-format entries also. Appendices 4.1 to 4.3 give a
demonstration of a few of the various types of indexes that could be
produced from a set of single DSIS input strings.

An alphanumerically arranged set of DSIS index entries having both Lead
and Context Headings give rise to what is called an 'Associative-cum-
Organizing' index. Consider the following set of alphabetically arranged
Lead Headings with Upper Link specifiers:

1 Maintenance skills, Personnel, Industries, Labour economics
2 Personnel, Academic institutions, Education
3 Personnel, Academic libraries, Librarianship
4 Personnel, Industries, Labour economics
5 Skills, Personnel, Industries, Labour economics

The Lead Headings numbered 1, 4 and 5 above are hierarchically related.
Although, the hierarchy is not ascertainable from the alphabetic
arrangement. But the alphabetical arrangement has brought together all the
three subject headings having 'Personnel' as the Lead Term, irrespective of
which hierarchy each one belongs. This grouping together of subject headings
having a common factor (in this case, alphabetical contiguity) without
reflecting hierarchy is an 'Associative Grouping'. Now, consider the
following Context Headings formed from the Lead Headings given above and
arranged alphanumerically:

1 Education 8 Academic institutions 8.4 Personnel 8.6 Non-teaching
   personnel 8.2 Skills 8.1 Improvement
2 Labour economics 8 Industries 8.4 Personnel
3 Labour economics 8 Industries 8.4 Personnel 8.2 Skills
4 Labour economics 8 Industries 8.4 Personnel 8.2 Skills 8.2.6
   Maintenance skills
5 Librarianship 8 Libraries 8.6 Academic libraries 8.4 Personnel

8.6 Clerical assistants 8.1 Training

The above sequence of Context Headings reflect an 'Organizing grouping'. Context Headings 2 to 4 reflect their hierarchy from broader to narrower subject representation in their vertical sequence, the intension of the subject represented being reflected by the horizontal length of the individual headings.

4.8 Vocabulary Control

It is evident from the methodology of DSIS (especially Section 3.2) that, it is necessary to have a standard tool for vocabulary control for this system of indexing. This tool can be designed either before starting the indexing work or along with the indexing work. Whatever may be the procedure, this tool for DSIS would require some special features of its own. It has to be a faceted systematic scheme for hierarchical (organizing) classification incorporating all the essential features of a conventional thesaurus (Bhattacharyya, 1982a, p. 140). In other words, it has to be a combination of faceted classification and thesaurus. Because of this reason, it has been called Classaurus.

It consists of a systematic part complemented by an alphabetical index part. Like any classification scheme, the systematic part displays hierarchical (organizing) relationship among terms (viz., broader, narrower, coordinate) in its schedules (see Appendix 4.4, Part 1, displaying a portion of Entity Schedule and Action Schedule). Like a faceted (category-based) classification scheme there are separate schedules for each of the 'facets' or Elementary Categories, viz., Discipline, Entity, Property, and Action; with their respective Species/Type, Parts, Constituents and Special Modifiers. Also there are separate schedules for Common Modifiers. Like any conventional thesaurus each of the terms in the hierarchic schedules is
enriched with synonyms, quasi-synonyms, and antonyms in extended senses. Also, like a thesaurus, any term is permitted to appear in as many hierarchies as may be appropriate (Devadason, 1985b, p. 16). But, no non-hierarchically (associatively) related terms are enumerated in a Classaurus, because the "task of showing what is non-hierarchically related to what, and how they are related is left to the care of the indexing procedure (in this case DSIS)" (Bhattacharyya, 1981b, p. 17). Due to this absence of associative relationships, the hierarchy of the terms in a Classaurus is shown by simple indentions. The purpose for which a Classaurus is used does not necessarily warrant any principle-based arrangement of the terms in an array. Even if the terms in each array are arranged alphabetically, the purpose is not going to be disturbed. This feature makes it largely suitable for computerization. By following this approach, an 'Alphabetic Classaurus' has been generated by Devadason and Kothanda Ramanujam (1982).

The Alphabetic Part is a chain index (see Appendix 4.4, Part 2) to all the terms in the Systematic Part. There is no index number in the alphabetic part. But each entry has at its end, the alphabetic code for the schedule to which the entry belongs. Each entry contains, for each and every term, its broader terms. Synonymous terms also act as access terms. To locate the position of a specific term in the systematic part, the sequence of terms in the chain entry in the alphabetic part is reversed. With the help of the alphabetic code for the schedule indicated at the end of each entry the appropriate schedule in the systematic part is consulted. The schedule is then searched successively using each of the terms in the reversed chain entry till the specific term is found. Once the specific term is located, then all the other terms in the hierarchy along with the synonymous terms could be reached.
4.8.1 Construction of Classaurus

There are three basic aspects in the construction of Classaurus, viz.,

1) selection of key terms or descriptors;
2) establishment of the ECs of key terms; and
3) establishment of interrelationships among the selected and categorized key terms as to whether they form broader or narrower or coordinate or synonymous relationship.

According to Devadason (1986a, p. 34), "'Classaurus' could be automatically created from names of subjects formulated and modulated as per DS of SIL, and kept on-line for referring to it and for keeping it always up-to-date".

4.9 Salient Features of DSIS

It has been claimed by Devadason (1986a, p. 33) that "The Deep Structure Indexing System is simple, built from quite simple constructs and operations". The three different types of terms, viz., 1) Compound Term, 2) Complex Term, and 3) Composite Term, comprise the basic building blocks of subject headings. Apart from formulating the subject proposition as per DS of SIL and indicating the ECs by appropriate codes, the indexer's work involves just the following three aspects:

1 Choice of Lead Terms, including those for PCR entries;
2 Choice of Context Terms; and
3 Formation of CR entries, due to vocabulary control.

The availability of default options makes the first two options still simpler and easier.

One of the salient features of DSIS is that, it has avoided the need for providing See also cross reference entries of the 'ascending' (from narrower to broader subjects) or "descending" (from broader to narrower subjects) types in general, when the Context Headings used are Full Context
Headings. Consider the Context Headings given under the Lead Term 'Basic education' in Appendix 4.3. All the Full Context Headings having this term are grouped together, which automatically brings together all the Context Headings having narrower terms to 'Basic education' also, such as 'Adult basic education', 'Literacy education', etc. If one searches 'Basic education' he need not be directed to search also under narrower terms to it, using See also entries of the descending type like PRECIS (Austin, 1984, p. 210). Similarly, the Full Context Headings given under the Lead Term 'Adult basic education', also have the broader term 'Basic education'. If the searcher wants information on broader terms to the term he is looking at, he gets those terms from the Full Context Heading under the term. Hence, there is no need for See also entries of ascending type. This approach results into reduction in the number of times a searcher has to look up a new expression to continue a search, i.e., lesser number of two-step searches. The process of creating an organizing classification effect in the verbal plane (i.e., by referencing) has been criticized as a case of "running from pillar to post" (Bhattacharyya, 1981a, pp. 98-99).

Computerized information retrieval from machine-readable databases using the same procedure may seem to be a solution to this problem, because of the speed and the least effort on the part of the searchers. But this solution is more apparent than real. Moreover, the rules of syntax of DSIS requiring the Discipline term to be represented as the first context specifying category eliminates the need for See also cross references of the 'associative' type 'Learning See also Education' and 'Education See also Learning', practised in PRECIS (Austin, 1984, p. 209). Attempts to enumerate non-hierarchically related terms of a particular manifestation, such as above, has only created confusions. No two specialists agree in the choice of so-called 'Related Terms'. But if this part of the activity is left to the care of the subject strings themselves, no two specialists can disagree; for, in this process two terms are said to be related because they have
occurred as related in the sources of information (Bhattacharyya, 1982b, p. 264). DSIS incorporates all these features.

Another notable feature of this system is that, by keeping the respective Full Context Headings as a key to each of the main entries (bibliographic references) and sorting them on this key, an 'organizing effect' can be produced in their sequence. When the main entries are printed in this sequence with renumbered entry numbers, along with their respective Full Context Headings as 'feature heading' printed on top of each main entry, then the subject index prepared for them need not necessarily be an 'Associative-cum-Organizing' one. An associative index using just the Lead Headings alone (together with the respective entry numbers) may be sufficient. On the other hand, if the main entries are not in the organizing sequence, then Associative-cum-Organizing index would be needed. In other words, DSIS offers a choice of different types of index entries to suit the need of the individual indexing situation concerned.

Unlike PRECIS, DSIS does not require a separate input procedure for the generation of its vocabulary control tool Classaurus, which could be created automatically from the input strings prepared according to DS of SIL. It could be used simultaneously and kept online for referring to it to prepare CR entries to control synonyms due to factoring of Composite Terms, and for 'modulating' and 'standardizing' the names of subjects. Keeping it always up-to-date is also possible (Devadason, 1985b). Also by augmenting the input names of subjects by a different set of codes, an alphabetical thesaurus could be generated automatically (Devadason and Balasubramanian, 1981; Devadason, 1983).

Devadason (1985a) has suggested that it is possible to use the Lead Terms and their associated Full Context Headings in an online information retrieval system. The system could be made to display the Full Context
Headings in which the search term occurs and upon selecting the relevant Context Headings, the bibliographic citations could be retrieved and displayed. If the number of Context Headings is beyond certain limits, one or more other search terms could be input to select Context Headings having a combination of them. Thus, we would have an easily searchable system with a built-in vocabulary control tool incorporated in it. It needs no prior knowledge of any query language for searching on the part of the searcher. Besides, the process codes and the category indicators used are simple and few in number, which make the life much easier for the programmer [8]. The computer can generate different types of subject index entries as well as the vocabulary control tool Classaurus using a single set of coded input strings.

Notes

1 However, there are doubts about the name, DSIS, as Devadason (1986b) himself says that "DSIS is a methodology and provides guidelines and not a 'hard and fast rules-based' system. My calling the computerised POPS3 as DSIS is not that correct". For the purpose of this study we stick to the name DSIS, though, frequent mention of POPS3I will be evident, as it is well-nigh impossible to delve into a study of the former without giving due reference to the latter.

2 "A statement or any other formulation in a language -- natural or artificial -- denoting a subject, is a Subject Proposition" (Bhattacharyya, 1979c, p. 15). For example,

Preservation of manuscripts in archives (according to natural language)
Preservation: Manuscripts: Archives (according to chain indexing)
Preservation. Manuscripts. Archives (according to PRECIS index)

In this study the expression 'Subject Proposition' has been used
3 The term 'manifestation' refers to a specific idea falling in any one of the Elementary Categories and also the term denoting it.

4 The Complex Term 'Libraries for Blind' cannot be formed as a Compound Term, say 'Blind libraries', for, it is not libraries which are blind but blind persons who are being served in those libraries. It should be noted here that, those Modifiers forming Compound Term in one natural language might require the insertion of auxiliary/function words in another natural language and vice versa. Moreover, that component in a subject proposition represented as a Complex Term (Modifier of Kind 1) is likely to be changed into a Compound Term (Modifier of Kind 2)/term by subsequent emergence of new technical terms in the subject field concerned. Therefore, if it is possible to represent a component in a subject proposition by both Modifiers of Kind 1 and Kind 2, the representation by Modifier of Kind 2 should be preferred in DSIS.

5 "Generally it is not necessary to 'modulate' Modifier of Kind 1 forming Complex Term. But if the Modifier of Kind 1 term occurs in the Classaurus for the concerned subject area as a manifestation of any one of the ECs: Entity, Property or Action (and not just a Modifier alone), then it may be worthwhile to include its broader terms also." (Devadason, 1986a, p. 10).

6 The process codes used to generate the Permuted Cross Reference entries are described in section 5.5 below.

7 According to Ranganathan's chain indexing system a 'chain' is "A modulated sequence of Subordinate Classes or Isolates (or Isolate Ideas)" (Ranganathan, 1964a, p. 285), and a 'link' is "A class (or Isolate Idea) in a chain" (Ranganathan, 1967b, p. 63).

8 However, as we shall see later, incorporation of the modifications proposed by this study will certainly leave the indexer with many additional decisions to make using extra process codes. Consequently, the
software has to make provisions to deal with them and will cease to be called simple.
References


Devadason, F. J. (1986b). Personal communication.

Devadason, F. J. and Balasubramanian, V. (1981). Computer generation of


Storage and Retrieval, 3 (4): 399-410.


APPENDIX 4.1

Uni-component Term Lead Heading and Full Context Heading

Adult basic education

Education 9.6 Basic education 9.6 Adult basic education 4 Africa == e0003

Adult education

Education 9.6 Adult education 0 Essays == e0001
Education 9.6 Adult education 4 Great Britain 4.4 England 0 Inquiry reports == e0004
Education 9.6 Adult education 8 Women 8.6 Working class women == e0006
Education 9.6 Adult education 9.2 Methodologies 9.2.6 Distance study 9.2.5 (using) Television services == e0009

Adult education (role of) Public libraries

Education 9.6 Adult education 9.5 (role of) Public libraries 4 Great Britain 4.4 England 4.4 Kent == e0007

Adult education (using) Computer systems (and) Interactive videos

Education 9.6 Adult education 9.5 (using) Computer systems 9.5 (and) Interactive videos 4 United States 0 Handbooks == e0005

Adult literacy education


Basic education

Education 9.6 Basic education 9.6 Adult basic education 4 Africa == e0003
Education 9.6 Basic education 9.6 Literacy education 4 Great Britain 3 Rural areas == e0002

Computer systems (and) Interactive videos/Adult education (using)

Digital computer systems = Computer systems
APPENDIX 4.2

Uni-component Term Lead Heading and Short Context Heading

Adult basic education

Education 9.6 Adult basic education 4 Africa == e0003

Adult education

Education 9.6 Adult education 0 Essays == e0001
Education 9.6 Adult education 4 England 0 Inquiry reports == e0004
Education 9.6 Adult education 8 Working class women == e0006
Education 9.6 Adult education 9.2 Distance study 9.2.5 (using)
  Television services == e0009

Adult education (role of) Public libraries

Education 9.6 Adult education 9.5 (role of) Public libraries 4 England
  4.4 Kent == e0007

Adult education (using) Computer systems (and) Interactive videos

Education 9.6 Adult education 9.5 (using) Computer systems 9.5 (and)
  Interactive videos 4 United States 0 Handbooks == e0005

Adult literacy education

Education 9.6 Adult literacy education 4 India 2 /1970-1980/ == e0010

Basic education

Education 9.6 Adult basic education 4 Africa == e0003
Education 9.6 Literacy education 4 Great Britain 3 Rural areas == e0002
Education 9.6 Adult literacy education 4 India 2 /1970-1980/ == e0010

Computer systems (and) Interactive videos/Adult education (using)

Digital computer systems = Computer systems

Distance study (using) Television services

Education 9.6 Adult education 9.2 Distance study 9.2.5 (using)
  Television services == e0009

Interactive videos/Adult education (using) Computer systems (and)
APPENDIX 4.3

Lead Heading with Upper Link Specifiers and Full Context Heading

Adult basic education, Education

Education 9.6 Basic education 9.6 Adult basic education 4 Africa == e0003

Adult education, Education

Education 9.6 Adult education 0 Essays == e0001

Education 9.6 Adult education 4 Great Britain 4.4 England 0 Inquiry reports == e0004

Education 9.6 Adult education 8 Women 8.6 Working class women == e0006

Education 9.6 Adult education 9.2 Methodologies 9.2.6 Distance study 9.2.5 (using) Television services == e0009

Adult education (role of) Public libraries, Education

Education 9.6 Adult education 9.5 (role of) Public libraries 4 Great Britain 4.4 England 4.4 Kent == e0007

Adult education (using) Computer systems (and) Interactive videos, Education

Education 9.6 Adult education 9.5 (using) Computer systems 9.5 (and) Interactive videos 4 United States 0 Handbooks == e0005

Adult literacy education, Basic education, Education


Basic education, Education

Education 9.6 Basic education 9.6 Adult basic education 4 Africa == e0003

Education 9.6 Basic education 9.6 Literacy education 4 Great Britain 3 Rural areas == e0002


Computer systems (and) Interactive videos/Adult education (using)

Digital computer systems = Computer systems
### APPENDIX 4.4

**Classaurus of Library Science Terms**

#### Part 1. Systematic Part

<table>
<thead>
<tr>
<th><strong>Action Schedule</strong></th>
<th><strong>Entity Schedule</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition</td>
<td>Libraries</td>
</tr>
<tr>
<td>.Accession</td>
<td>(for)</td>
</tr>
<tr>
<td>.Ordering</td>
<td>-Blind</td>
</tr>
<tr>
<td>.Selection</td>
<td>-Mentally handicapped</td>
</tr>
<tr>
<td>Administration</td>
<td>-Seafarers</td>
</tr>
<tr>
<td>Circulation</td>
<td>Academic libraries</td>
</tr>
<tr>
<td>Technical processing</td>
<td>=Libraries for educational institutions</td>
</tr>
<tr>
<td>.Cataloguing</td>
<td></td>
</tr>
<tr>
<td>(in)</td>
<td>.College libraries</td>
</tr>
<tr>
<td>-Publication</td>
<td>.School libraries</td>
</tr>
<tr>
<td>=CIP</td>
<td>...Elementary school libraries</td>
</tr>
<tr>
<td>..Author cataloguing</td>
<td>...Secondary school libraries</td>
</tr>
<tr>
<td>..Computerized cataloguing</td>
<td>.University libraries</td>
</tr>
<tr>
<td>..Descriptive cataloguing</td>
<td>.National libraries</td>
</tr>
<tr>
<td>.Indexing</td>
<td>.Public libraries</td>
</tr>
<tr>
<td>.Subject indexing</td>
<td>.Municipal libraries</td>
</tr>
<tr>
<td>..Computerized subject indexing</td>
<td>.Rural libraries</td>
</tr>
<tr>
<td>.Subject classification</td>
<td>.Special libraries</td>
</tr>
<tr>
<td>..Online subject classification</td>
<td>.Research libraries</td>
</tr>
</tbody>
</table>
Part 2. Alphabetic Part

Academic libraries 8.6 Libraries (E)
Accession .1.4 Acquisition (A)
Acquisition (A)
Administration (A)
Author cataloguing .1.6 Cataloguing .1.4 Technical processing (A)
Blind (for) 8.5 Libraries (E)
Cataloguing .1.4 Technical processing (A)
CIP = Publication (in) .1.5 Cataloguing
College libraries 8.6 Academic libraries 8.6 Libraries (E)
Computerized cataloguing .1.6 Cataloguing .1.4 Technical processing (A)
Computerized subject indexing .1.6 Subject indexing .1.6 Indexing .1.4
Technical processing (A)
Descriptive cataloguing .1.6 Cataloguing .1.4 Technical processing (A)
Elementary school libraries 8.6 School libraries 8.6 Academic libraries 8.6 Libraries (E)
Indexing .1.4 Technical processing (A)
Libraries (E)
Libraries for educational institutions = Academic libraries
Mentally handicapped (for) 8.5 Libraries (E)
Municipal libraries 8.6 Public libraries 8.6 Libraries (E)
Online subject classification .1.6 Subject classification .1.4 Technical processing (A)
Ordering .1.4 Acquisition (A)
Publication (in) .1.5 Cataloguing .1.4 Technical processing (A)
Research libraries 8.6 Special libraries 8.6 Libraries (E)
Rural libraries 8.6 Public Libraries 8.6 Libraries (E)
School libraries 8.6 Academic libraries 8.6 Libraries (E)
Seafarers (for) 8.5 Libraries (E)
Secondary school libraries 8.6 School libraries 8.6 Academic libraries 8.6 Libraries (E)

Selection .1.4 Acquisition (A)

Special libraries 8.6 Libraries (E)

Subject classification .1.4 Technical processing (A)

Subject indexing .1.6 Indexing (A)

Technical processing (A)

University libraries 8.6 Academic libraries 8.6 Libraries (E)
Although it has scientific foundations, bibliographical control is primarily a technology. The word 'technology' has its origin in the Greek word 'techne', meaning 'skill' or 'art'. The skill or art involved in the practice of bibliographical control is that of organizing information for retrieval purposes. As a result, research carried out in this field is basically of 'how-to-do-it' type. This is of two kinds: (1) problem solving or developmental research, e.g., how to computerize DSIS; and (2) decision making or evaluative research, e.g., is PRECIS more efficient than DSIS or vice versa (Svenonious, 1981, p. 88). Thus, the former relates to the design of new systems, while the latter results, as we have exemplified, in the comparison of two systems, which again have implications for the design of new systems.

5.1 Index Language Comparisons

Bibliographical control includes indexing in its entirety, of which subject indexing is a sub-field. An obvious evaluative research question in the area of subject indexing is which of the several systems for indicating the subject of a body of information is superior, should it be alphabetic or classed, pre-coordinate or post-coordinate, controlled or free-text, manual or automated, and so on. Again, among any one of these types, say pre-coordinate systems, should it be any one of those existing such as, LCSH, PRECIS, POPSI, etc., or a completely new system yet to be designed. Although the question is more than a century old, there is a radical change in what constitutes an acceptable answer to it, especially since the time of Cranfield experiment I (Cleverdon, 1962). The experiment, though suffering from technical drawbacks, brought to light the fact that it is possible to
evaluate the performance of indexing systems on the basis of objective
criteria such as precision and recall, rather than on subjective judgements
as practiced till the late 1950s.

The popularity of Cranfield I resulted in a spurt of new experiments. The major and representative ones among these are characteristically represented by Sparck Jones (1981) in an excellent historical review spanning the period 1958-1978. Based on the shift of emphasis in the aspects of retrieval systems on which the tests were conducted, the period was divided into two decades, viz., 1958-1968 and 1968-1978, respectively. Within each individual period the tests were broadly categorized into: (a) indexing language tests, and (b) automatic indexing tests, with the exception that the last decade was divided into two further categories, viz., indexing tests and searching tests. The number of tests conducted during these two decades cannot be said to be insignificant, if one takes into account the age of information science as an independent discipline. But little attempt has been made to build up new experiments based on the findings of the tests conducted in the past. As a result, information retrieval experiments as a sub-field within the broader spectrum of information science, lack in both the existence of a sound methodology and general hypothesis-building. It has been pointed out that, most of such tests suffer from poor design, exhibiting lack of experimental controls, lack of external validity (due to limited test data and artificial search requests), and lack of validity in the definition of independent and dependent variables. Some of these could be attributed, of course, to the practical side of the tests, i.e., the nature of the data (documents and user queries) and the mechanism (indexing and searching) involved. It is still very difficult to find out what constitutes an 'ideal test collection'. On the mechanism side, so far emphasis has been on the indexing operation, the searching half (shall we say 'better half'?) being more or
less neglected. Only recently, information scientists have began to realize its importance in the context of the whole retrieval system. Also, there has been a tendency to move away from the comparison of total systems toward singling out certain individual features of index languages (such as, synthetic capability, degree of specificity, coextensivity, consistency, etc.) and seeing how variations in these affect user satisfaction (Swanson, 1966; Svenonious, 1971; Markey, 1984). In the area of machine or machine-assisted indexing, attempts were made to improve the techniques of automatic extraction of descriptors or keywords from text, which relates as much to searching as to indexing them (Liebesny, 1974; Sparck Jones, 1974). Despite all these shortcomings and shift of emphasis, it has been confirmed that, at least some substantial contributions have been made by the retrieval system experiments. The tests have shown:

"(1) that artificial indexing languages do not perform strikingly better than natural language;
(2) that complex structured descriptions do not perform strikingly better than simple ones;
(3) that the number of searching keys is more important than their individual quality;
(4) that the characterization of queries is more important than that of documents;
(5) that formal properties of the data may be turned to advantage, as in weighting schemes" (Sparck Jones, 1981, p. 248).

As far as both PRECIS and DSIS is concerned, the first two statements go straightway against them, because both belong to the category of artificial index languages, and also, as we have described them in succession in chapters 3 and 4, respectively, both use complex and highly structured descriptions to represent the thought content of a document. The rest of the findings follow as a corollary to the first two, which one way
or other indicate that these two systems stand positively on the other side of the fence. Artificial index languages are 'artificial', because they employ some form of control on their structure and function, as against the natural languages which impose no control (in the 'de jure' sense). Therefore, artificial index languages can be labelled 'controlled languages'; this control may be applied both at the input (indexing) and/or output (searching) stages. This control can take the form of hierarchy as in classification schedules and thesauri) and/or simple synonym-homonym control (as in the keyword indexes), both essentially ending up with the establishment of classes of terms. Terms, apart from being related at the level of hierarchy and equivalence, can also be related in some other way of association. Controlled index languages use various types of relationship between classes of terms to represent the subject of documents, as depicted in the order of books on shelves or cards in a catalogue or in a file of index entries (both online and printed). These relations may be expressed by the provision of syntactical devices in an index language, which range from facet indicators (as in faceted classification), and roles and links (as in post-coordinate systems) to most recent relational operators (as in Relational Indexing, PRECIS). Application of these devices establishes in varying degrees of accuracy the relationships between the elemental components in the name of a subject. The height of accuracy is claimed for schemes of relational operators, such as those devised by Farradane and by Austin. PRECIS's twin process of categorical analysis (i.e., establishment of 'deep cases' through role operators) and relational analysis (i.e., writing of the input strings through recognition of certain primitive relationships) of concepts are seen as comparable in a way to the technique of facet analysis, though with limited classificatory effect (Vickery, 1975, p. 106). In this sense, we can call PRECIS a classificatory index language. On the other hand, the effect and use of classificatory principles are quite
explicit in DSIS.

In the light of the index language evaluations carried out during the 1960s and the 1970s, Keen (1976) assessed the performance and potential merits or faults of classificatory index languages, and arrived at the following conclusions which apply to PRECIS and DSIS (in prospect) among others:

(1) different types of classificatory index language do not substantially differ in performance merit;

(2) non-classificatory index languages do not have a substantially different performance from classificatory ones;

(3) formal hierarchical linkage has not been found to be essential to high recall;

(4) syntactical devices as precision improvers have a small and minority value; and

(5) the index language, as one of several sub-systems in a complete information retrieval system, is of minor importance.

To many this might seem to be enough to drive a final nail in the coffin of controlled index languages, especially those based on classificatory principles. But, the real situation appears to be quite different. Not only the old classificatory index languages have survived (for example, Dewey Decimal Classification is running into its 20th edition with major expansions and renewed popularity), but a host of others have proliferated, the base of which is certainly classification (Aitchison, 1986; Aitchison and Gilchrist, 1987, p. 79-100). In fact, in the USA (the proclaimed land of non-traditional information retrieval), Batty (1981) has evidenced a renewed interest in the use of facet analysis, the time-honoured tool of classification, in a number of index languages, such as ERIC, BIOSIS, etc. So it seems, Keen’s (1976, p. 156) apprehension had some validity, when he predicted that the 1980s would "see the protagonists of particular kinds of
index language sticking to their guns in spite of the evidence". But the fact is, after a decade of deliberation the so-called protagonists seem not to have lost their heart and have looked for support elsewhere, e.g., psychology, linguists, artificial intelligence, etc. The future will decide, how much of these will pass the test of time and be of practical use for the ever benevolent user of systems using these languages, and how much of these will remain an exercise in academia slowly passing into oblivion. Meanwhile, in the background, the debate between controlled versus uncontrolled languages is continuing (Austin, 1986; Borko, 1986; Svenonius, 1986; Balasubramanian, 1988). Nevertheless, we must acknowledge considerable improvements as a result of greater knowledge of index language structures and functions and as a result of increasing experience of an evergrowing variety of indexing languages in an ever widening range of environments and applications. In this respect, the contributions of evaluation tests and the insights gained from such tests are nonetheless important. However, with a paperless society in the offing (Lancaster, 1978), doubts have been raised about the whole issue of bibliographic control such as, 'do we need indexing at all' (Svenonious, 1981, p. 93). But at the same time, apart from many new index languages being developed, there seems to be no dearth of such effort as to compare the efficiency and effectiveness of competing indexing systems (Wellisch, 1980, 1984), although the credibility of the methodology adopted and the validity of the hypothesis drawn are never beyond doubt.

5.2 Comparisons between PRECIS and POPS

Only two specific tests involving PRECIS could be named truly evaluative by nature. These are Keen's (1978) EPSILON (Evaluation of Printed Subject Indexes by Laboratory investigation) project and the Wollongong University Subject Catalogue Study (WUSCS) (Hunt et al, 1976/77). The former was done
in a laboratory environment, while the latter in an operational setting. With formatting kept constant, and input strings more or less so, the EPSILON investigators tested the efficacy of printed subject indexes constructed by schemes such as chain procedure, PRECIS, articulated, and rotated (KWAC). The data constituted a subset of the ISILT (Information Science Index Language Test) (Keen and Digger, 1972) test collection. During the test, it was realized for the first time that clarity of presentation of information could be an effective criterion in judging most information retrieval systems. This was found to be of special importance in a printed index. A variety of evaluation methods were used, including a questionnaire to determine searchers' preferences. Some findings of this test which bear direct relevance to string indexing systems such as PRECIS and DSIS, were:

1. Preservation of entry context allows significant rejection of non-relevant entries for very little recall loss.

2. The varieties of function word provision and term order (e.g., PRECIS) perform indistinguishably (Keen, 1981).

The main limitations about EPSILON are restriction to a single subject area and the small size of its database. The WUSCS project also evaluated string indexing systems such as PRECIS, LCSH (Library of Congress Subject Headings) and KWOC (all with slight variations), primarily from the point of view of output. The test data included a little more than 2000 documents spread over eight disciplines. There was hardly any significant advantage in one system over the others both from the input as well as output point of view. Here again the major limitation noted was the relatively small size of the database, which combined with the variety of disciplines, tended to bias the results against the more sophisticated systems, such as PRECIS. On the other hand, DSIS (or its predecessor POPSIL) is yet to appear in any test of truely evaluative proportion.
So far as our interests are concerned, very few comparative studies have been carried out between PRECIS and POPSI (not to mention of DSIS), of which the works of Mahapatra (1978), Rajan (1976), Rajan and Guha (1979), Farradane (1977), and Bhattacharjee (1981), are the notable ones. After studying the syntactical differences between POPSI and PRECIS, Mahapatra (1978) recommended that the two systems should not be considered as alternative to one another, since they cannot replace each other, but can develop as independent yet interacting systems. Rajan (1976) described the handling of compound terms in three major pre-coordinate indexing systems, viz., POPSI, BTI and PRECIS, and made a comparative assessment of their approaches to the problem with the final suggestion that the classificatory approach (explicit in POPSI, but implicit in BTI) would perhaps be more helpful than the linguistic approach (as in PRECIS). Among these, the comparative study by Rajan and Guha (1979) can be considered the most extensive one though unfortunately now out of date. Many of the features of PRECIS they criticized had been changed long before this work was published. It included comparisons of PRECIS and POPSI strings and chains, format and display of entries, syntax, etc, and discussed their suitability for machine handling and consistency. It was finally observed that the ultimate solution lies in adopting more classificatory ideas for indexing purposes, and recent developments in indexing systems have brought us much nearer to a neutral and universally acceptable syntax. Farradane (1977) compared computer-produced alphabetical subject indexes by the methods of PRECIS, NEPHIS, an adaptation of Relational Indexing and POPSI with the help of 12 examples and emphasized the desirability of information retrieval tests of these systems. More recent attempt (Bhattacharjee, 1981) to study PRECIS and POPSI together was confined to a comparison of their basic postulates and subject heading structuring. In another study, Dutta and Sinha (1984) made a survey of existing indexing systems such as SLIC (Selected Listing In Combination), PRECIS and POPSI, for the production of computerized subject indexes at the
Sorghum and Millets Information Center (SMIC), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) Library, India. But decided to go for a more 'Pragmatic Approach to Subject Indexing' (PASI) than the postulational approaches advocated by the existing systems mentioned above. POSFI was specifically ruled out due to the absence of a readily available Classaurus in the subject field concerned. Craven (1986) has proposed a common and effective technique for weighing choices among indexing systems, which involves: i) construction of a table of choices and features, ii) the rating of each choice on each feature, iii) the weighing of the features, iv) the multiplication of the ratings by the corresponding feature weights, and v) the summing of the resulting products to calculate an overall value for each choice. Based on this technique he made a tentative comparison among seven string indexing systems, viz., PERMUTERM (of Institute for Scientific Information), ABC-Spindex (of ABC-Clio), PRECIS, POSFI, CIFT (Contextual Indexing and Faceted Taxonomic Access System), CASIN (Computer Aided Subject INdex), and NEPHIS (NEested Phrase Indexing System). POSFI (along with PERMUTERM, ABC-Spindex) was rated much lower than any of the runners-up (PRECIS, CIFT), not to mention the winners (CASIN, NEPHIS).

5.3 The Present Comparative Study

5.3.1 Objectives in General

We have already mentioned in the introductory chapter (Chap.1) that, the purpose of this study is to make a comparison between PRECIS and DSIS, and to study the latter's possibility of replacing PRECIS in future and its contribution to our quest for an optimal indexing system. Carrying out a full-fledged investigation, needs both resources and suitable operational environment. Neither of these were available under the present circumstances. The other alternative way was to carry out a laboratory test
under controlled environment. But the absence of a ready-made DSIS index forced to abandon this possibility. Next attempt was to write an index string generator for DSIS that can produce a standard DSIS index from a sample collection of documents. But lack of programming ability again proved a deterrent. The software written is yet to perform satisfactorily. Some of the problems we faced specifically are mentioned in the section entitled 'EXHIBITS' (No. III), appended at the end of the thesis. Finally, it was decided to carry out an evaluation, based mainly on theoretical and hypothetical premises. Apart from the rigour and experimental control needed to measure the variables in manual testing at the searching stage, this less than hundred percent satisfactory DSIS index, also made it pointless to conduct a test of the searchers' preferences. Keen (1976, p. 142) has identified three main types of activities which can be described as index language evaluations:

(a) those involving examination of particular classification or other index language, in isolation from any documents and search requests;
(b) those involving examination and comparison of one or more index languages by classifying or indexing particular set of documents, again divorced from an actual index or the operation of searching; and
(c) those taking the form of testing a complete information retrieval system, whether in the controlled environment of a laboratory experiment or the real-world situation of an operating system.

Therefore, this comparative study can be labelled as belonging to the second category.

Two particular concepts included in the title of this thesis need some explanation. These are the concepts of 'efficiency' and 'effectiveness', respectively. There are many different aspects or properties of a system that one might wish to measure or observe, but majority of them are concerned with the effectiveness of the system, or its benefits, or its
efficiency. In simple terms, effectiveness is how well the system does what it is supposed to do and its efficiency is how cheaply it does what it does. The debate on what comprises the effectiveness of an information retrieval system is long and involved. Important contributions have been made by Cleverdon (1972), Lancaster (1979), Van Rijsbergen (1979), etc. For example, search effectiveness is often measured by dividing retrieved items into those relevant to a given query and those not relevant. In other words, "it is a measure of the ability of the system to retrieve relevant documents while at the same time holding back non-relevant ones" (Van Rijsbergen, 1979, p. 145). Two common measures of effectiveness that make use of this division are the recall and precision ratios. The 'recall ratio' is the proportion of relevant documents retrieved among all relevant documents; the 'precision ratio', the proportion of relevant documents retrieved among all retrieved documents. It should be noted that, no effort has been made to quantify these measures of effectiveness, although some indication is given in the form of observations and predictions.

The other criterion for evaluation, viz., efficiency, considers variables such as time, cost, cost/benefit, cost/effectiveness. Whether online or printed, a major advantage which string indexing systems such as PRECIS, DSIS, etc., have over other kinds of systems is that they can be generated with a relatively small investment of input data. A number of entries can be generated automatically from a single input string plus its address. Apart from reducing the work of the indexers and others involved at the input stage, such systems can make some savings on the costs of data storage and transmission, as well. The present study looked into these matters, specifically the relative ease and economy of input preparation, each system's capability in handling subjects of various denomination and complexity, absence of redundant and irrelevant information in the respective index entries, and the relative bulk of the indexes
produced. Again it must be emphasized that no quantification has been attempted. Also, the systems were studied against the various international standards related to the field of concept analysis, subject indexing and thesaurus construction. Both the indexes and vocabularies (used and generated) were tested in terms of both linguistic and classificatory principles.

From the point of view of searching, string index displays hold certain distinct advantages over other types of indexes. In general, searches for information will be more effective the more relatively useful information and the less relatively useless information the searchers find. Searches in index displays will be more efficient the less effort searchers have to spend to achieve the same amount of effectiveness. In fact, "the main purpose of the multiple overlapping index entries and explicit syntactic rules characteristic of string indexes is to improve the efficiency and effectiveness of searches" (Craven, 1986, pp. 6-7). Besides efficiency and effectiveness, other factors which to some extent affect the searches are, as mentioned above, redundant and irrelevant information, and index bulk. The relative performance of these factors related to searching are covered in a separate chapter (Chap. 8).

5.3.2 Materials and Input Preparation

Three sets of two hundred documents each, equally divided into macro and micro, respectively related to the subject fields of adult education, information retrieval and labour economics, were recorded on "Input Record Sheets" from secondary sources. In the case of macro documents such as books, monographs, etc. the only secondary source being used was the British National Bibliography (London: The British Library, 1951- ). For the micro documents such as journal articles, research papers, etc. sources were respectively Current Index to Journals in Education (Pheonix, Arizona, USA;
The first step in any indexing process involves writing out expressions such as "this book is about...", "this article is about...", etc. To aid in writing such an indicative formulation that summarises in its message "what a particular body of information is about", or what is simply known as a 'title-like phrase', the title of the document was supplemented by additional terms selected from feature headings (in the case of macro documents), abstracts (for micro documents), etc. The procedures followed are those recommended in the British Standard 6529 (British Standards Institution, 1984) and its equivalent International Standard for 'examining documents, determining their subjects and selecting indexing terms'. Two examples are provided here to make this point intelligible:

(1) Clarification of the Original Title basing on the Feature Headings:

[Cross, Michael. Towards the flexible craftsman. London: Technological Change Centre, c1985]


*BNB Dewey Decimal Classification broad class no.:* 331 - Labour economics.

*Expressive Title:* In labour economics, effects of technological change on the maintenance skills of personnel in engineering industries.

*[Note: Of course, for PRECIS strings there is no need to include the name of the discipline term.]*

(2) Inclusion of New Terms from the Abstract:

Abstract: Discusses Library of Congress shortcomings in the subject treatment of Jewish materials suggesting that it falls short of the goals of access and equity. Outlines several aspects of the vocabulary problem and its application which has been illustrated with real cases. Suggests 2 ways of improving subject cataloguing.

Expressive Title: In information retrieval, treatment of Jewish materials in Library of Congress Subject Headings scheme.

Next, PRECIS and DSIS input strings for each of the six hundred items were constructed based on the steps and procedures recommended in the Manual (Austin, 1984) and FID/CR report (Devadason, 1986a), respectively. It has been claimed by Devadason that, altogether fourteen different types of subject index entries (which include twelve types of POPSI entries plus a pseudo PRECIS-format and chain index entries) could be automatically created from a single coded DSIS input string, out of which any one could be selected as appropriate, depending on the purpose at hand. However, for the purpose of this study, we have decided in favour of the type "Lead Heading with Upper Link specifiers and Full Context Heading" (see Appendix, Chap. 4), as it reveals the full implication of DSIS principles and procedures in the index generated. The BNB Subject Authority Fiche (London: The British Library, September 1985) with minor modifications was used as the standard vocabulary control tool for both the systems. The reasoning behind this decision was the desire to keep terminological variations to a minimum level as between the PRECIS and DSIS indexes. For example, care has been taken to avoid the use of certain terms like 'Information retrieval systems using computer systems', which BNB/PRECIS uses as a single concept. But for both PRECIS and DSIS input string writing purposes this has been changed into its acceptable compound form in the English language as 'Computerized
information retrieval systems' and input respectively as,

*(1) information systems $2$1 computerized

and

8 $0$ Information systems 8.6 $3$ Computerized

These will be rendered as

Information systems

Computerized information systems

and

Information systems

8 Information systems 8.6 Computerized information systems

Computerized information systems

8 Information systems 8.6 Computerized information systems

While preparing the input strings care has been taken to avoid any bias in the construction of indexes. But despite all efforts certain amount of bias could not be avoided in the indexes, especially in the categorization of action/property concepts. For example, terms such as 'Economic relations', 'diseases' would figure in the action category in PRECIS, because such terms represent what Austin calls 'phenomena', which appear to represent things engaged in action rather than an action per se and at the same time these cannot be reduced to infinitives. But according to Ranganathan's Colon Classification, these terms would belong to the Matter-Property category. DSIS also adheres to this theory in principle. But, to avoid certain awkward constructions resulting from the use of Modifier of Kind 1 concepts (explained in successive chapters), it was decided to categorize such concepts as action concepts rather than as properties. To many this might seem an undue bias (for example, Devadason (1986b) cautioned at the beginning of this project by saying "one should unlearn PRECIS in order to
learn DSIS"), which, to a lesser extent, could be attributed to the fact that, the same person was involved in the production of both the indexes. It is practically impossible to be totally unbiased. Experience has shown that, even rigorously controlled laboratory tests were prone to such undue influences exerted by the construction of one index over the others (Keen, 1981, p. 142). Such influences can also be observed in the choice of vocabulary, particularly in choosing the BNB Subject Authority Fiche as the standard vocabulary control tool which reflects the practice followed in a certain environment and culture. The project used the in-house online PRECIS index system available for teaching and research in the department (Smith, 1986). The DSIS software is in its laboratory stage and subject to further tests. Both programs are written in CBASIC (of Digital Research) and run on a Comart microcomputer system. The format of indexes used for comparison is mainly printed with occasional reference to their online versions.
References


Devadason, F. J. (1986b). Personal communication.


CHAPTER 6
SUBJECT HEADING STRUCTURING

Like all other language systems, concepts in subject indexing languages, such as PRECIS and DSIS, are basically related on two main axes, viz., semantic and syntactic, respectively. We have already seen that, this division is drawn on the basis of certain pragmatic factors rather than on any solid theoretical foundations. For the sake of convenience, Austin has tried to have separate discussions on each of them in different sections of the *Manual*. But in certain cases (e.g., in the case of differencing operators) the discussions overlap with each other. In the case of DSIS, such separation becomes still more difficult, because the rules of the system call for the incorporation of both types of relations into the entry structure. The additional structure, viz., 'elementary structure' performs as an intermediary through which syntactic and semantic structures are knitted together. Here it becomes more problematic to discuss them in isolation. Whether PRECIS or DSIS, in an operational environment, however, the indexer performs both the operations almost simultaneously. Despite all these practical obstacles, it has been decided to have separate discussions on each category of relationship. This chapter is basically devoted to the study and comparison of the syntactical aspects of both the systems, concentrating on the handling of compound terms, syntax, format and display of index entries.

6.1 Treatment of Compound Terms

It would be useful to remember at the very beginning that, this is one of the areas where semantic and syntactic relations largely overlap with each other. But, nonetheless, the discussion of term or concept structure is essential, in the sense that terms constitute the very basic unit in the
name of a subject. However, an attempt will be made in this chapter to isolate its implications on the vocabulary from that of its grammar and leave the discussion of the former till the next chapter.

A 'term' (indexing term) is the verbal representation of a concept, preferably in the form of a noun or noun phrase, which an indexer ascribes to a document to describe its subject matter (International Organization for Standardization, 1986; subsequently referred as ISO 2788). A simple concept such as 'Libraries' is presented by a term consisting of only one noun word. But, concepts such as 'National libraries' and 'Libraries for children' are represented by compound terms consisting of noun phrases with more than one word. According to the standard, a 'compound term' is an indexing term which can be factored morphologically into separate components, each of which could be expressed, or reexpressed, as a noun that is capable of serving independently as an indexing term. The parts of most compound terms can be distinguished into:

(a) the 'focus' (or 'head'), i.e., the noun component which identifies the broader class of concepts to which the term as a whole refers, e.g., the noun component 'Libraries' in both the phrases above; and

(b) the 'difference' (or 'modifier'), i.e., one or more further components which serve to narrow the extension of the focus and so specify one of its subclasses, e.g., the adjective 'National' in the compound term 'National libraries', the preposition-plus-noun combination 'for children' in the compound term 'Libraries for children'.

Based on the type of difference attached to the focal noun, noun phrases belonging to the category of compound terms can be classed into:

(a) Adjectival phrases, e.g., 'National libraries'. This class also includes those single-word compounds which can be factored morphologically into a
noun plus a modifying difference having an adjectival function, e.g., 'Videodiscs'.

(b) prepositional phrases, e.g., 'Libraries for children'.

PRECIS follows this criterion. But according to DSIS, these two types will be treated separately as compound term (formed by modifier of kind 2) and complex term (formed by modifier of kind 1), respectively.

6.1.1 Adjectival Compounds

Both the systems perform similarly when the noun is modified by preceding adjectives which can be specified as first and second level differences (dependent and independent modifiers in DSIS). For example, a subject like 'Mentally handicapped readers' will be rendered as:

PRECIS: *(1) readers $21 handicapped $22 mentally

DSIS: 8 readers 8.6 $3 handicapped 8.6 $3 mentally

and will generate the natural left-to-right order:

Figure 6.1: Treatment of Adjectival Compounds (Type I): PRECIS/DSIS

Here each word modifies the next component of the phrase. Changing the order of adjectives would make it meaningless, as in 'Handicapped mentally readers'. Problems arise only when a focus ('modifyee' in DSIS) is qualified by two or more differences on the same level (i.e., two or more independent modifiers); for example when each of two adjectives is related directly to the noun, neither adjective modifying the other. For example, a subject like 'Fee-based online bibliographic information systems' will be input in PRECIS
as:

*(1) information systems $21 bibliographic $21 online $21 fee-based

and will also generate the natural left-to-right order:

```
1st level differences -- focus
fee-based online bibliographic information systems
```

Figure 6.2: Treatment of Adjectival Compounds (Type II): PRECIS

The DSIS input string will be:

```
8 information systems 8.6 $3 bibliographic 8.6 $3 online 8.6 $3 fee-based
```

and will also generate the same natural left-to-right order, but in a different manner:

```
dep. modifier    dep. modifier    indep. modifier    modifyee
fee-based        online           bibliographic      information systems
```

Figure 6.3: Treatment of Adjectival Compounds (Type II): DSIS

Consequently, this produces a basic difference. The above PRECIS input string will generate the following lead headings with the whole of the term as their display or context:

- Information systems
- Bibliographic information systems
- Online information systems
- Fee-based information systems

because each of the above modifiers represent a different characteristic of
division (such as content, medium, and cost). Whereas, the DSIS input string will generate the following lead headings followed by the same block of terms forming the context:

Information systems
Bibliographic information systems
Online bibliographic information systems
Fee-based online bibliographic information systems

But the fact is, there may be other documents represented by lead headings such as 'Online information systems', 'Fee-based information systems', etc. It should be noted that someone interested in either of these two subjects will also find it useful to consult the document dealing with the subject mentioned above. But the search for 'Online information systems' or 'Fee-based information systems' will not easily retrieve 'Fee-based online bibliographic information systems'. Due to alphabetic adjacency, the search may retrieve it close to the entry 'Fee-based information systems', but not to the entry 'Online information systems'. Because the latter will be far down in the alphabetical order, especially in a large file such as LISA (Library and Information Science Abstracts). This may well result in loss of relevant information and considerable reduction in the recall value of the system (Biswa and Smith, 1988a, p. 8). Thus we can say, the collocation in a PRECIS heading is better than the DSIS heading. Only when the search begins from the focal term 'Information systems' as lead, will the recall be hundred percent.

Bhattacharyya (1982, p. 251) suggests that "there will always be the need to permutation [sic] in such a situation". Though he has not made it clear whether these permutations are to be entered in the index in the form of subject index entries with the permuted terms as Lead Headings and Context Headings, or as Cross Reference entries parallel to the type used
for controlling naturally occurring synonyms, quasi-synonyms, etc. But the outcome will be obviously more than one input string for a single subject statement, which is self-defeating for the purpose of string indexing. It can be observed that, total reliance on the natural language order (of the author) and the Wall-Picture Principle would give birth to multiple unconnected hierarchies and resulting loss of collocation among index entries. As a result, the index would fail to perform one of the three functions of an index entry, viz., the 'relating function', which is being defined as "the location of entries for topics related to the one being sought" (Keen, 1977a, p. 19). It is true that "Indexing as a process in which we are involved, is document oriented indexing" and "The relation between index terms should be based on (are brought out by) the individual document being indexed and on the subject (area) treated in the document" (Devadason, 1986b). But it is also true that it is the information (idea) contained in the document which we are interested in at the end, not the document per se. The above situation makes it apparent that certain additional procedures must be introduced to standardize and control the use of authors' description of subjects for indexing purpose. Our aim should be to satisfy 'Every user (reader) his/her information (book)' supplemented by 'Every information (book) its user (reader)'.

6.1.2 Prepositional Phrases

In English, as in other languages, compound concepts are sometimes expressed in the form of prepositional phrases, such as 'Libraries for children', 'Schools for blind', etc. In English language grammar this is designated as the postmodified noun phrase, which could be effectively changed into its premodified form of adjectival phrase (e.g., 'Children's libraries'). Devadason (1986a, p. 3) asserted that "modifiers forming Compound Term (i.e., adjectival compounds) in one natural language may require the
insertion of auxiliary/function words and form a Complex Term (i.e., prepositional phrase) in another natural language and vice versa...
Moreover, depending on the usage of terms in a subject area, that component in the name of a subject represented as a Complex Term is likely to be changed into a Compound Term/term by subsequent emergence of new technical terms". Hence, if a component in the name of a subject admits of representation by both adjectival compounds and prepositional phrase, the former is preferred in the DSIS. However, recognised standards such as ISO 2788 or its equivalent BS 5723 (British Standards Institution, 1987) did not lay down guidelines which enable an indexer to determine whether a concept should be expressed in its adjectival form or as a prepositional phrase. According to Austin (1974, p. 87), there is very little logical difference between a sub-class of a focus which is specified adjectivally and another which is specified by means of a prepositional phrase. Prepositional phrases, like the adjectival constructions can be analysed into a focus and difference, e.g.

\[
\begin{array}{c}
\text{focus} \quad \text{---} \quad \text{difference} \\
\text{Libraries} \quad \text{for children}
\end{array}
\]

Figure 6.4: Treatment of Prepositional Phrases: PRECIS/DSIS

The solution suggested in PRECIS is that in a limited number of cases the indexer can apply the technique used for adjectival differences to terms containing these following differences, i.e., by coding the substantive which follows the preposition as though it was the focal term. For example,

*(1) children $21 libraries for

This will produce the following entries:

Children

Libraries for children
Libraries for children

However, this procedure must be done with caution and the noun which follows the preposition should not be selected as a lead "unless it can share the same context (s) as the focus in any circumstances". Logically, this sort of treatment can be labelled unsound. For example, it led Coates (1976, pp. 92, 95-96) to remark "What can one say of an indexing system for subject analysis which lays claim to a logical basis, but which on occasion 'twists' the meaning of its ostensibly meaningful relational operators?" and "its handling of prepositional relationships is to an unacceptable extent arbitrary and too dependent upon linguistic accidents". However, Austin's general recommendation is to input such terms in their natural language order and provide access to the differencing term (s) through a See also reference extracted from the thesaurus, e.g.

Children

See also

Libraries for children

Welfare services for children

But this solution is also not without its limits. It can be quoted as one of those fatal attempts to include purely physically every type of relation between terms in the thesauri (Fugmann, 1974, p.77).

Compared to this, the solution offered by Devadason is rather straightforward. The input string for the above topic will be:

8 <libraries 8.5 $* (for)/ $2 children>

which will generate the following entries:

Libraries (for) Children

Libraries 8.5 (for) Children
The above treatment is superior to PRECIS for the following reasons: a) there is no need to 'twist' the meaning of the category indicators at the input stage; and b) the second entry is not a main entry, it is a permuted cross reference entry, directing the searcher to look under 'Libraries (for) children' and removing the possibility of any initial misinterpretation.

6.1.3 Complex Subjects

In DSIS complex subjects formed using auxiliary/function words are also categorised along with the multiple-word terms using prepositions in them (Devadason, 1986a, p. 4). A "Subject formed by coupling two or more subjects expounding, or on the basis of, some relation between them" constitutes a complex subject (Ranganathan, 1967b, p. 85) and is formed by using phase relations, such as comparison, bias, influence, etc., e.g., 'Statistical methods biased to librarians', 'Term entry indexing systems compared with item entry indexing systems', etc. Ranganathan introduced five kinds of phase relation -- General, Bias, Comparison, Difference and Influence. In addition to the above five, Bhattacharyya (1979, p. 18) prescribed two more, viz., Similarity and Application. But "Complex Subjects formed by phase relations are generally narrower than the subject represented by the first phase" (Devadason, 1986a, p. 4) seems to be an over generalisation. In the examples cited above 'Statistical methods biased to librarians' is narrower than 'Statistical methods' in general and can be considered as a species/type, but it is difficult to accept that 'Term entry indexing systems compared to item entry indexing systems' is narrower than 'Term entry indexing systems'. Also, on certain occasions terms created by phase relations in them, such as 'Indexing using computer systems' may be later amenable to form 'fused subjects' represented by adjectival compounds like...
'Computerized indexing' it is difficult to foresee any such development in technical terminology which can transform complex subjects such as 'Term entry indexing systems compared to item entry indexing systems' into an acceptable adjectival compound term. It seems too much of an attempt to fit every case into a single straightjacket. The factoring of compound terms in PRECIS conforms to the rules formulated in international standards (ISO 2788 and/or BS 5723) with few additions. The BS 5723 suggests that "Complex subjects should be expressed by combinations of separate terms, and these may be assigned as independent search keys in a post-coordinate system, or they may function as components of pre-coordinated index entries. It is realized, however, that this general recommendation does not stipulate the exact circumstances in which a compound term encountered in documents should be factored into separate components" (italics mine) (BS 5723, p. 9).

Further, it is proposed by Devadason (1986a, p. 10) that, "Generally it is not necessary to 'modulate' Modifier of Kind 1 forming Complex Term. But if the Modifier of Kind 1 term occurs in the Classaurus for the concerned subject area as a manifestation of any of the ECs: Entity, Property, or Action (and not just a Modifier alone), then it may be worthwhile to include its broader terms also". But how? No suggestions have been put forward as to the manner in which such Modifier of Kind 1 terms are to be modulated. Certainly it cannot be incorporated within the Complex Term block enclosed within the angular brackets. For example, given that a section of the Entity schedule in the Classaurus for 'Library and information science' is,

Information systems
  .Information processing systems
    ..Data processing systems
      ...Computer systems
        ....Expert systems
and the subject to be indexed is 'Evaluation of medical information retrieval using expert systems'. This could be analysed and formalised as follows:


Now, if we modulate Modifier of Kind 1 term by augmenting it by interpolating the successive superordinates, then we would get:

(D) Library and information science, (E) Information, (Type of E) Medical information, (A) Retrieval (m1) (using) Information systems, (Type of m1) Information processing systems, (Type of m1) Data processing systems, (Type of m1) Computer systems, (Type of m1) Expert systems, (A on A) Evaluation

[Note: m1 = Modifier of Kind 1]

After replacing auxiliary words denoting the different manifestations with appropriate indicators, the resulting name of subject would be:

Library and information science 8 Information 8.6 Medical information 8.1 Retrieval 8.1.5 (using) Information systems 8.1.6 Information processing systems 8.1.6 Data processing systems 8.1.6 Computer systems 8.1.6 Expert systems 8.1.1 Evaluation

So far, so good. But if one considers the practicalities of searching such renderings, one might conclude that "Introduction of... superordinate links in a subject string where they are superfluous leads to confusion among information system users and therefore possible misinterpretation of the subject strings" (Raghavan and Iyer, 1978, p. 12). As in this case, by the time the user reaches 'Evaluation' he could easily have lost the link
between 'Retrieval' (modifyee or focus) and 'Expert systems' (Modifier). He may by now be thinking that it is an 'Evaluation of expert systems for... (something)', rather than 'Evaluation of medical information retrieval using expert systems'. This seems to be the most serious drawback of modulating Modifier of Kind 1 terms forming part of the Complex Term, even if it appears in the Classaurus as a manifestation of one of the Elementary Categories. Of course, the same could be said to be true for the whole system as it proposes to modulate the subject statement by interpolating and/or extrapolating as the case may be, the successive superordinates of each EC manifestation by finding out of which it is a Species/Type or Part or Constituent. This practice certainly leads to an enormous increase in the number of terms in a subject string. While users' surveys have found out that "Subject strings with eight or more component terms present difficulties in interpretation" (Raghavan and Iyer, 1978, p. 12). A similar problem in relation to PRECIS' long headings and their overrun onto a second line, was pointed out by Keen (1977b).

6.2 Syntax

Two separate but occasionally overlapping trends can be visualized in the formulation of indexing language syntax. The first, providing a 'categorical' view of syntactic structures in indexing, has its origin in traditional classification systems, which despite the improvements offered by Ranganathan's theories of facet analysis and Vickery's extention of facet categories beyond PMEST, viewed the process in terms only of inclusion relation, of hierarchies in which subject-disciplines or concepts are by definition contained by other subject-disciplines or concepts, as in a series of 'nested' boxes (Coates, 1973, p. 392). POPSI as well as DSIS belong to this family of classificatory syntax indexing systems, since their syntax is determined according to the Postulates and Principles of Facet
Analysis as expounded by Ranganathan (1967b). But by the beginning of the second half of the present century it was realized that a facet structure defined exclusively in terms of categories of concepts occasionally failed to ensure unambiguity of meaning. Also the facets are themselves essentially the superficial manifestations of relations between concepts belonging to different facet categories. Farradane's (1950) system of indexing based on relational analysis could be considered as a breakaway from the classificatory syntax tradition, which gives more importance to the meaning of the spaces between constituent terms in an index string to characterize its syntactic structure. Gardin's (1965) SYNTOL and Coates' (1970) BTI also have great merit in discriminating relations, but also have recourse to categories. PRECIS is characterized as the most elaborate relation-based syntactic indexing system developed for operational use during 1960s, although some of its 'roles' can be defined in categorical terms (Coates, 1978, p. 294). Before debating the relative merits of the two syntaxes, let us first examine what constitutes the syntax of a string indexing system. Two separate mechanisms are available for symbolising syntactical relations within an index string. They are, first, the order in which terms appear in it, alternatively known as the citation order, and second, the connectives used to indicate links between terms, i.e., relational connectors. The citation order of terms in a PRECIS string follows the principle of 'context dependency'. That is, the order of the elements in a string of words describing a subject is arranged so that each term is dependent upon the term before it and sets the context for the term following it. When terms have been organised in this way, they frequently form what is called a 'one-to-one related sequence' (Austin, 1984, pp. 10-11). The 'shunting' technique helps to generate this sequence. Compared to this, the citation order of terms in a DSIS string follows an alleged 'deep structure' of subject indexing languages developed by Bhattacharyya, which in essence matches the theory of single-entry citation order of Ranaganathan, who suggested that
there may be an 'absolute syntax' for subject descriptions (Ranganathan, 1967a). Consistent application of only one citation order yields better predictability in the index entries. However, in a more complex situation it may fail to provide desirable results. For example, if terms denoting Common Modifiers (in DSIS) or extra-core operators (in PRECIS) are marked to form lead, PRECIS strings gives better collocation by putting more significant terms more immediately after the lead term, e.g.,

**PRECIS:** Bibliographies

Children. Diseases
Spain. Political events
Women. Attitudes

**DSIS:** Bibliographies, Attitudes, Women, Sociology

Sociology 8 Women 8.2 Attitudes 0 Bibliographies
Bibliographies, Diseases, Child medicine, Medicine
Medicine 9.6 Child medicine 8.2 Diseases 0 Bibliographies
Bibliographies, Political events, Spain, History
History 8 Spain 8.2 Political events 0 Bibliographies

However, in a DSIS index without upper link specifiers in the lead heading, the output becomes more or less same as that of PRECIS, e.g.,

**Bibliographies**

History 8 Spain 8.2 Political events 0 Bibliographies
Medicine 9.6 Child medicine 8.2 Diseases 0 Bibliographies
Sociology 8 Women 8.2 Attitudes 0 Bibliographies

Our experience has shown that, at least on one occasion DSIS fails to generate logically viable index string and predictable index entries. For example, we have a name of subject 'Curriculum for in-service training of non-teaching personnel in polytechnics', which is indexed in PRECIS as:
It can be noticed that, the term 'curriculum' representing some sort of indirect agent/instrument and coded 3 is preceded by a blank field coded 2. This does not prevent the functioning of the 'predicate transformation', but the computer, following instructions, assigns the empty field to the display position. The rest of the terms are printed in the qualifier position. In effect, the blank field identifies the space that would have been occupied by the implied action for which 'curriculum' is responsible, which might be expressed by a phrase such as 'implementation'. On the other hand, in DSIS, as we shall see, the existing rules of syntax fail to provide a meaningful input string. In DSIS the rules of syntax give rise to a context-dependent sequence of the components in the name of subject in conformity with Ranganathan's Principles for Facet Sequence -- the Wall-Picture Principle and its derivatives such as the Actand-Action-Actor-Tool Principle.

The 'Actand-Action-Actor-Tool Principle' has been defined by Ranganathan (1967b, p. 428) as, "If in a subject, facet B denotes action on facet A by facet C, with facet D as the tool, then the four facets should be arranged in the sequence A, B, C, D". For example, we have a name of subject in our hand: 'In education, curriculum for in-service training of non-teaching personnel in polytechnics'. Here the Action is 'In-service training'; the
Actant is 'Non-teaching personnel'; the Actor is absent, but implied (maybe teachers or tutors); and the Tool (instrument or agent) is 'Curriculum'. The sequence between 'Polytechnics' and 'Non-teaching personnel' is determined by the 'Whole-Organ Principle', which says "If, in a subject, facet "B" is an organ of facet "A", then A should precede B" (Ranganathan, 1967b, p. 427). Therefore, we shall have 'Polytechnics. Non-teaching personnel', the latter being an organ or part of the former. Of course, everything will be preceded by the Discipline term 'Education'. Therefore, when expressed in transformed skeleton form, we shall have 'Education. Polytechnics. Non-teaching personnel. In-service training. Curriculum'. This result can also be achieved by the repeated application of the Wall-Picture Principle. The above name of subject could be analysed and formalised according to DSIS as:

(Discipline) Education, (Entity) Polytechnics, (Part of Entity) Non-teaching personnel, (Action on Part of Entity) In-service training, (Tool or Instrument or Agent) Curriculum

After applying EC indicators we get:

Education 8 Polytechnics 8.4 Non-teaching personnel 8.1 In-service training [*] Curriculum

If one takes a closer look he will find that in both the expressions above the term 'Curriculum' has not been coded following DSIS procedures (indicated by the "[*]"). Because the system does not provide any suitable EC for it in the above name of subject. The nearest possible solution may be to treat it as a Special Modifier. But again, it does not fall within the circumference of the structure of the 'Modified Term' deemed to form either Modifier of Kind 1 or Modifier of Kind 2. The former needs the insertion of suitable auxiliary/function words (in between) to form an acceptable natural language title-like phrase, which is not possible in this case; while the latter requires to form an acceptable Compound Term automatically, which it
also fails to be. Let us see what Devadason has to say in this respect. In his opinion, the above document says very little about 'Polytechnics', it is more about 'Curriculums or syllabi or courses of study'. So do we. But the same could be said about the document on 'In leather technology, dry salt curing of pig skin using drums', which is input as:

Leather technology 8 Hide and skin 8.4 Skin 8.6 Pig skin 8.1 Beamhouse operation 8.1.4 Curing 8.1.6 Salt curing 8.1.6 Dry salt curing 8.1.5 (using) Drum

The above document definitely gives more importance to the 'method of preservation (of pig skin)' rather than the 'pig skin' itself. But it is the rules of syntax which determined the sequence of terms in this name of subject, not their relative importance. According to Devadason (1986b), "The UNMODULATED but formalised statement of the name of subject without much attention being paid for the sequence of modifiers (not much bothered about Wall-Picture principle) would be something like this (italics mine):

Education 8 Curriculum 8.5 (for) In-service training 8.5 (of) Non-teaching personnel 8.5 (in) Polytechnics".

This seems to deny those very basic principles upon which the whole system is built. In a similar study on 'Concept specification by PRECIS role operators' Mahapatra and Biswas (1985, p. 65) have found that a well-established and institutionalised system like PRECIS is also guilty of such 'manipulations'. The same conclusion could be put forward here that "This sort of input strings can be achieved by the indexer only when his mind is conditioned beforehand to somehow bring the required order of concepts to the index entries, and not the obvious relationships of concepts within the document, which might lead to poor results in the future". Leaving aside the obvious loss of clarity arising out of modulating all three .5 terms, such
practice also leads to a less predictable index string. Though the rules of syntax of DSIS exercise the main control over citation order while the discipline controls aspects important for the differences between disciplines, there must be a single citation order within a particular discipline so that predictability of the representation of concepts and concept relations is assured. Say, for example, another document minus the 'Curriculum' concept would definitely be input as:

Education 8 Polytechnics 8.4 Non-teaching personnel 8.1 In-service training
rather than as,

Education 8.1 In-service training 8.1.5 (of) Non-teaching personnel 8.1.5 (in) Polytechnics

Further, if the decision is to produce an index with short context headings (i.e., without the superordinate links), then the entry under 'Curriculum' carries an amount of redundancy, such as:

Curriculum (for) In-service training (of) Non-teaching personnel (in) Polytechnics, Education

Education 8 Curriculum 8.5 (for) In-service training 8.5 (of) Non-teaching personnel 8.5 (in) Polytechnics

It can be noticed that both lead and context headings are rendered in the same way, not reverse rendering (in lead heading) followed by forward rendering (in the context heading) (Devadason, 1986a, p. 20). Nonetheless, such entries increase the bulk of the index without any contribution to the search efficiency.

The main function of connectives in an index string, apart from separating terms, is to increase clarity or detail. Connectives may be of various kinds, according to whether we wish to maximize clarity of index
citation or minimize noise in searching. The main types of connectives are words and punctuation. PRECIS uses both. However, in addition to these two, a third type, viz., numerals, is used by DSIS. It is true that an alphanumeric arrangement improves collocation over a pure alphabetic order, but it considerably reduces the readability of the entries by increasing noise in searching. Although as Bhattacharyya (1981b, pp. 14-15) says that "the ordinal values of numerals do not need to be memorised. ...(and) the users are not concerned with these numerals (and) they can afford to ignore them completely", the numerals certainly do increase the 'weariness of the eyes'. Also for a first time user the burden of learning these conventions is considerable. One wonders, how little we think about the comfort of the poor user, who always comes out as the second best (the first being the information or document) in majority of information retrieval systems?

6.2.1 Main Class or Discipline as the First Category

We have already mentioned (Section 2.3.3.2, Chap. 2), the events leading to CRG's abandonment of the concept of main class as the basis for a library classification and acceptance of 'Universe of concepts', rather than 'Universe of subjects', for description of the components of the name of subject in all fields of knowledge. PRECIS was a by-product of this line of thinking. Coupled with the problems of chain indexing, especially the problems of disappearing chain and main class, BNB decided to have a verbal subject index for its printed bibliography free of any such main class constraint. In the later development of chain indexing, POPSI solved the problem of disappearing chain, but retained the discipline term as the first context specifying category. Devadason has claimed that, the inclusion of Discipline/Base as the first 'context specifying category' eliminates the need for See also cross references of the type 'Animals See also Zoology' and 'Zoology See also Animals' practised in PRECIS. But the feeling of the
present author is that it creates more problem than it solves. Firstly, a number of so-called discipline terms can function as action terms as well. For example, 'Education' is a discipline in the sense that, it is being taught and researched as a subject in educational institutions. Hence, it has been treated in most of the bibliographic classification schemes (such as, DDC, BC) as a main class. Also, it is an action on the ground that its origin can be traced to the verb 'to educate'. It creates the problem of citation order in the string, especially when terms denoting parts and types of education appear in the entry. For example, the term 'education' can further be modified by adjectives such as, 'vocational', 'technical', 'continuing', etc., as well as prepositional and auxiliary phrases such as 'using interactive videos', 'role of volunteers', etc., and form subclasses of 'education'. The implication for this in DSIS is that this will lead to a large array of sub-disciplines (appearing as large phrases, similar to the example of 'curriculums' in Section 3.3, above) relegating the more concrete entity terms further away from the beginning of the entry. Another negative impact of this idea of putting Discipline as the first context specifying category, pointed out by Biswas and Smith (1988a, p. 10), is that this very factor makes it unsuitable for a multi-disciplinary index such as that of BNB subject index. In DSIS (as well as in POPSI) the rules of syntax exercise the main control over the citation order while the Discipline controls aspect important for the differences between disciplines. For instance, consider the following treatment of the subject of a document 'Hunting of seals by Inuit', quoted from Craven (1986, p. 108). If the Discipline is marine biology, the terms may be cited in the order:

Marine biology 8 Seals 8.1 Hunting 8.1.5 (by) Inuit

If, on the other hand, the Discipline is anthropology, the order may be:

Anthropology 8 Inuit 8.2 Hunting 8.2.5 (of) Seals
Though, this idea of limited control of citation order by the Discipline of the indexed items takes some account of needs of different searchers for different kinds of access, nonetheless, it consumes more input time and output space, hence, in general would be uneconomic to produce.

6.2.2 Coordinate Relationship

The majority of the subject propositions are made up of concepts related in a strictly one-to-one fashion, leading to a linear structure in the input strings. However, many documents deal with concepts which share a coordinate relationship with some other component in the name of subject. This coordinate relationship calls for the addition of a second dimension to the linear structure. In ordinary language, parallel or coordinate parts of descriptors are usually marked by the presence of the conjunction, such as 'and'. For example, a subject like, 'Installation of microfilm readers and photocopiers in public libraries' could be represented as,

![Diagram](image)

Figure 6.5: Levels of Coordination

In describing their methodology of facet structuring of subjects for the identification of non-hierarchical associative relationships (NHR) [1] among ideas forming components of subjects, Neelameghan and Maitra (1977, p. 9) have also confirmed that "every type of NHR can be represented by one or other of the following relations -- facet relation, speciator relation, phase relation and coordinate relation -- in the facet analysed representation of subjects" (italics mine). The 'coordinate relation' is defined as the 'relation between two or more ideas in one and the same
array, derived from a broader or superordinate idea on the basis of a single characteristic for the division'.

In a situation such as the above, the PRECIS input string indicates the first term in the 'coordinate block' (i.e., microfilm readers and photocopiers) by a primary or dependent element operator and the other coordinate term(s) is written as the next component in the string and preceded by the operator g, e.g.,

String: *(1) public libraries
   *(p) microfilm readers $v &
   *(g) photocopiers
(2) installation

Entries: Public Libraries
   Microfilm readers & photocopiers. Installation
   Microfilm readers. Public libraries
   Installation
   Photocopiers. Public libraries
   Installation

POPSI (as well as DSIS) prescribes that "what is non-hierarchically related to what, will be revealed by the subject-propositions themselves through their alphabetical arrangement. ...for, in this process two terms are said to be related because they have occurred as related in the sources of information" (Bhattacharyya, 1981a, p. 101). Therefore, all the associative relationships (i.e., NHRs), including the coordinate relationship, are to be revealed through the index entry itself, not through any RT-type (or, See also) cross-references from the thesaurus (in fact, Classaurus does not include them at all). So it is obvious that there would be some provision for the treatment of such coordinate terms in the system. But DSIS hardly gives any consideration to documents dealing with such coordinate concepts, except in 'special situations' where "the conjunction 'and' may also be used
in a Complex Term to form a multifocal component" (Devadason, 1986a, p. 15), such as:

Information retrieval 8.1.5 (using) Interactive videos 8.1.5 (and)

Microcomputers

According to Devadason "multifocal or multi-theme documents would require separate names of subject for each theme". It seems that Devadason has confused multitopic (multi-theme) works and multielement (multifocal) works (cf. Angell, 1972, p. 152). "A multitopic work treats a number of discrete subjects, each of which may be denoted specifically by a separate index term. A multielement work, on the other hand, treats of subjects which are not discrete but are so intimately bound together that a string of interconnected terms is required to denote the work specifically" (Svenonius and Schmierer, 1977, p. 338). So far as multi-theme documents are concerned this treatment is satisfactory, but may not be feasible for documents dealing with coordinate concepts. The failure lies in the added work required of the indexer and of the index string generator. Besides the increase in the amount of input work (apart from those 34 strings involving the 'special situation' mentioned above, there were 51 input strings in the sample containing one or more such coordinate concepts) on several occasions it may lead to loss of intelligibility. For example, if a subject like 'Integration of managerial and technical skills of middle-managers in commercial banks' is coded separately as,

Commercial banks 8.4 Middle-managers 8.2 Managerial skills

8.2.1 Integration

and

Commercial banks 8.4 Middle-managers 8.2 Technical skills

8.2.1 Integration

then the last term in both the strings become misleading, since the concept 'Integration' logically refers to both 'Managerial skills' and 'Technical
skills' as a whole unit, not to each individual preceding term separately. Austin (1984, p. 97) would designate them as 'Bound coordinate concepts' and replace the operator g with operator f, which gives greater latitude and less redundancy in an "apples and oranges" situation, and at the same time improving the grammatical readability of the entries (Richmond, 1981, 115). For example,

String: ' (1) commercial banks
(p) middle-managers
*(p) skills $21 managerial $v&
(f) skills $21 technical
*(2) integration

Entries: Skills. Middle-managers. Commercial banks
Managerial skills & technical skills. Integration
Managerial skills. Middle-managers. Commercial banks
Managerial skills & technical skills. Integration
Integration. Managerial skills & technical skills.
Middle-managers. Commercial banks

6.2.3 Viewpoint and Target

There are times in specifying the subject of a document when it may become necessary to indicate the viewpoint from which the subject is examined, for the author's viewpoint can significantly affect the user's perception of relevance. In Langridge's (1986, p. 225) opinion, "Fundamental disciplines are by far the most important formal characteristic of documents, but there appear to be six other categories, apart from physical features, with varying degrees of significance". One of these six he refers to as 'viewpoint'. Ranganathan even showed awareness of its importance by making viewpoint the primary facet (base) in certain main classes in his Colon classification, the system facet in medicine and psychology being examples.
According to Austin (1974, p. 220), such viewpoint terms can be regarded as a special kind of inner form, which does not match our normal understanding of the form of a document, viz., physical or intellectual (as 'serials', 'biographies', 'atlases', etc.), and target (as 'for librarianship', 'for trade unions', etc). Another important distinction is that terms designated as form refer to all the preceding concepts in the string, whereas the viewpoint relates more directly to the core concepts in the subject-proposition. It is, therefore, necessary to introduce this concept at a position in the string closer to the terms prefixed by the main operators. For instance, consider the following subject: 'Publishers' viewpoint of copyright infringement by libraries in Great Britain during 1970-1985'. In this subject the 'Publishers' viewpoint' directly relates to the issue of 'Copyright infringement by libraries', whereas 'Great Britain' and '1970-1985' merely add the place and time dimensions to it. PRECIS designates such viewpoint/perspective/aspect terms as 'extra-core factors', which is introduced by operator 4 and input as follows:

(0) Great Britain
*(1) copyright
(2) infringement $d 1970-1985 $v by $w of
*(3) libraries
(4) publishers' viewpoints

and will generate the following entries:

Copyright. Great Britain

Infringement by libraries, 1970-1985 -- Publishers' viewpoints

Libraries. Great Britain

Infringement of copyright, 1970-1985 -- Publishers' viewpoints

In their comparative study of PRECIS and POPS1, Rajan and Guha (1979, p. 379) did not find any similar provision in POPS1. But in a later study,
Bhattacherjee (1981, p. 132) showed that concepts denoting 'Viewpoint-as-form' (operator 4) in PRECIS can be designated as 'Speciators' in POPS1. For example, a subject such as 'In economics, evaluation of industrial relations from trade union point of view' can give rise to the following subject heading in POPS1:

Economics (BS) ; Industrial relations (MP) ; Evaluation - Trade union viewpoint (E)

[where BS = Basic Subject or Discipline; MP = Matter-Property or Property in DSIS; E = Energy or Action; "-" = Indicator for Speciator]

Bhattacharyya (1981b, p. 14) calls it 'Special Modifier', rather than 'Speciator'. But this treatment is also unsatisfactory on the ground that, instead of being treated as a Common Modifier, it has been input as a Special Modifier. Whatever one calls it, the foregoing discussion shows that there is a genuine need to make provision for introducing such viewpoint terms in indexes, which DSIS certainly fails to perform. A similar problem can be experienced in the treatment of 'Target-as-a-form' concepts, which is also represented as Modifier of Kind 1 in DSIS, especially when the subject has more than one isolate, that is to say, a compound subject.

6.3 Entry Format

In comparison with PRECIS' use of three different entry formats, viz., standard format, predicate transformation and inverted format, DSIS uses only one format for the entries. In the case of PRECIS, the need for 'predicate transformation' was envisaged due to the fact that, occasionally, the position of a term is not of its own accord sufficient to indicate its role beyond reasonable doubt. Such a situation arises specifically when two or more entity terms are related to an action term, variously representing different deep cases such as, 'Object', 'Instrument', 'Agent', etc. The use
of the third format, viz., 'inverted format' is more of a practical convenience, in the sense that, it tends to improve collocation by putting more significant terms more immediately after the lead term; any effect caused by a sacrifice of the context principle is considered relatively mild. On the other hand, in order to avoid any possible confusion as that arising from the the need to apply 'predicate transformation' in PRECIS, the sequence of component terms in Context Headings is kept invariant in DSIS, along with the different indicators. It is claimed that such a fixed order provides better comprehension and resolves any ambiguity that may arise. But this regularity can be seen as rigidity. Despite the possibility that this order can be varied according to the purpose of the index and the need of the users, in practical terms it is more difficult to apply, especially in a general and multi-purpose index.

6.3.1 Choice of Lead Term

So far as the choice of the lead term in an index is concerned, it is well known that significant or sought terms are usually marked to form lead, but not diffuse or heavily-used terms. Despite the fact that it is very difficult to ascertain which terms are significant and which are not, general decisions concerning leads are to be formulated within an organisation and recorded as policy statements. For example, very generic entity terms such as, man, animals, plants, etc., very common action terms such as, evaluation, analysis, determination, measurement, etc., very common property terms such as, efficiency, property, effectiveness, etc., and terms denoting forms, place, time, etc., need not be necessarily selected to form leads. Policies of this kind may differ from one organisation to another depending on the purpose of the subject index -- the subject field concerned, the material being indexed and the community of users served by the index. In PRECIS, whether or not a term should appear in the lead is to
be determined entirely by the indexer, not by the system or the computer (Austin, 1984, p. 24). It prescribes a "✓" (tick) to mark those terms under which the indexer wishes to generate leads. For example,

**String:**
(1) books
(2) production

**Entry:**
Books
Production

DSIS also adheres to the same prescription in general, except on one occasion. In the case of permuted cross reference (PCR) entries, which are formed by cyclic permutation of constituents in a complex term (i.e., multi-word terms using auxiliary/function words in them), the indexer is compelled to mark the first constituent of such term as lead even if he does not wish to do so. For instance, a string like

\$9 \text{Labour economics} 8 \text{Industries} 8.2 \text{productivity} 8.2.1 \langle \text{Measurement} 8.1.5 \text{* (by)}/ \text{$2 \text{Gross National Product ratio}\rangle\n
will generate the following entries:

- Industries, Labour economics
  - Labour economics 8 Industries 8.2 Productivity 8.2.1 Measurement
    - 8.2.5 (by) Gross National Product ratio c005
  - Productivity, Industries, Labour economics
    - Labour economics 8 Industries 8.2 Productivity 8.2.1 Measurement
      - 8.2.5 (by) Gross National Product ratio c005
  - Measurement (by) Gross National Product ratio, Productivity,
    - Industries, Labour economics
      - Labour economics 8 Industries 8.2 Productivity 8.2.1 Measurement
        - 8.2.5 (by) Gross National Product ratio c005
  - Gross National Product ratio / Measurement (by)
Now if the indexer decides not to have a lead on 'Measurement' but only on 'Gross National Product ratio', thinking that the former is too general a term to be sought by the users of the index, he is in a fix. Because there will not be any entry having the standard rendering of the complex term as lead (the third entry above), under which other sections such as the context and location appear. Therefore, the entry under 'Gross National Product ratio' (which is a PCR entry, having neither the context nor the location) virtually leads to a blank and become meaningless. In DSIS, to have a lead under the second or successive significant constituents of a complex term one must have a lead on the first constituent term irrespective of its status. Although, DSIS does not include See also cross references in general and 'organising' search can be performed directly through any one of the 'lead-only' terms, our practical experience has shown that the number of such PCR entries can be quite large and nonetheless prove irritating to the searcher.

Another important drawback of DSIS is the illogical subordination of the entries generated under the second or subsequent constituents of a complex term. This is particularly evident in complex terms involving some sort of interrelation, e.g., a 'two-way interaction' or an 'author-attributed association'. Consider the following subject involving reciprocal interactions: "Foreign relations between Great Britain and Soviet Russia". The DSIS string will be

$9 History 8 Great Britain 8.2 <Foreign relations 8.2.5 (with)/ $2 Soviet Russia>

which will create the following lead headings:

Great Britain, History

Foreign relations (with) Soviet Russia, Great Britain, History

Soviet Russia / Foreign relations (with)
As we can see there is no main entry under 'Soviet Russia'; due to the rules for forming permuted cross reference entries it has been relegated to the subordinate position of a cross reference entry. Such treatment might serve the purpose of patrons in a Russian library, who usually look under the name of other nations, but not every user everywhere interested in historical literature. Besides, there is a possibility that the searcher might misinterpret the second heading which is not very clear. The only way out seems to be to write a duplicate string inverting the sequence of components in the complex term (as in 'Soviet Russia foreign relations with Great Britain'). PRECIS solves this problem by using operator 'u' as follows:

String: *(1) Great Britain
      *(u) foreign relations $v with $w with
      *(1) Soviet Russia

Entries: Great Britain
         Foreign relations with Soviet Russia
         Foreign relations. Great Britain
         With Soviet Russia
         Foreign relations. Soviet Russia
         With Great Britain
         Soviet Russia
         Foreign relations with Great Britain

Although PRECIS produces more entries and thus increases index bulk, it does better justice to the subject by giving equal weight to both phases of the string. DSIS faces the same problem in indexing subjects involving some sort of author-attributed association, e.g., "Leninism compared with Maoism". In fact, the very structure of the complex term in DSIS forces the indexer to subordinate all Kind 1 Modifier terms from the modifyee (or focus). In the case of prepositional phrases, e.g., 'Libraries for blind', the implication of such subordination may not be very striking, in the sense that the
difference acts as a minor characteristic to form a subclass of the whole class of 'Libraries'. But seemingly on all other cases such subordination can hardly be justified.

6.3.2 Selection and Rendering of the Context Term

One can easily glimpse the kind of difficulties that could result from DSIS' attempt to repeat the entire input string, with all its generic terms, in every index entry. For example, a document on 'Computerized subject indexing of in-house journals in small research libraries' has the input string:

Library science 8 Libraries 8.6 Special libraries 8.6 Research libraries 8.6 Small research libraries 8.4 Documents 8.6 Serials 8.6 Journals 8.6 In-house journals 8.1 Documentation 8.1.6 Indexing 8.1.6 Subject indexing 8.1.6 Computerized subject indexing

and a searcher looking up 'Computerized subject indexing' encounters the index entry

Computerized subject indexing, In-house journals, Small research libraries, Library science

Library science 8 Libraries 8.6 Special libraries 8.6 Research libraries 8.6 Small research libraries 8.4 Documents 8.6 Serials 8.6 Journals 8.6 In-house journals 8.1 Documentation 8.1.6 Indexing 8.1.6 Subject indexing 8.1.6 Computerized subject indexing

Indeed, no matter from which direction searchers approach the document, they are always led down the complete generic-to-specific route:

Libraries -> ... -> Documents -> ... -> Computerized subject indexing

Following such a route seems overly time-consuming and the greater index bulk produced by the longer entries a hindrance to searchers. For the indexer, this seems to be an overly time-consuming and tedious process, since, for every near-similar subject statement he has to write the whole
hierarchy in the string. PRECIS usually excludes these extra terms from the index entries. They are handled by *See* and *See also* references extracted from a machine-held thesaurus. This exclusion of semantically related terms from subject statements is also, of course, a feature of natural language. Whenever we speak of say 'Journals' we do not need to state explicitly that we mean some sort of 'Documents', nor do we need to refer to 'Documentation' if we are talking about 'Subject indexing'. In each case, one of the terms (i.e., 'Documents' and 'Documentation') is present by implication as part of our normal frames of reference. In PRECIS, the general rule is that the name of a class or genus should not be included in a string that deals with its species if these concepts are linked by the true generic relationship. However, this general rule is sometimes broken. The circumstances which call for such action are, when the relationship between the genus and its species is not truly generic one, i.e., 'quasi-generic' (e.g., 'Paraprofessionals as library personnel'), the term is a homograph (e.g., 'Cranes' as 'Birds' and as 'Lifting equipment'), the term is a proper name representing a 'class-of-one' (e.g., 'Gandhi' as a 'Writer'). In such cases, PRECIS recommends the use of operator q, as shown below:

**String:** *(1) libraries
   (p) personnel
   *(q) paraprofessionals

**Entries:** Libraries
   Personnel: Paraprofessionals
   Paraprofessionals. Personnel. Libraries

Apart from these, this operator is also used for some practical purpose, e.g., to increase collocation under the broader term in certain entries. One question often asked by antagonists of PRECIS is 'how many links (are necessary) in string for context'. This is followed by apprehensions relating to the consistency of such practice and by conclusions, such as,
from a classificatory point of view, the above practice of PRECIS is something like naming the facet or pseudo-isolate first and then the specific isolate (Rajan and Guha, 1979, p. 373). The answer to the question is that, operational PRECIS indexes have rarely used more than one extra term to establish the context of a particular term. The apprehension expressed has nothing to do with PRECIS as a system, it solely depends on the policy formulation of the particular organization. The conclusions drawn are quite right, which have been pointed by Austin as one of the areas where semantics and syntax overlap with each other.

6.3.3 Types of Entries

Devadason (1986a, p. 32) has claimed that four major types of POPSI entries could be formed by the way the Lead Headings and Context Headings are formed in DSIS, viz.,

1) Uni component term Lead Heading with Full Context Heading;
2) Uni component term Lead Heading with Short Context Heading;
3) Lead Heading with Upper Link Specifiers and Full Context Heading; and
4) Lead Heading with Upper Link Specifiers and Short Context Heading.

Examples of subject index entries of types 1, 2, and 3 were provided in Exhibits 4, 5, and 6, respectively (Devadason, 1986a, pp. 39-41; also see, Appendices 4.1 to 4.3, Chap. 4 in this thesis). However, if one takes a closer look into the portion of subject index entries in Appendix 4.2, then it becomes evident that certain Lead Headings look dubious and out of context, e.g.,

Basic education

Education 9.6 Literacy education 4 Great Britain 3 Rural areas ==e0002

In the above example, certainly the user will find it difficult to establish
the proper context of the Lead Heading 'Basic education' as to whether it is a narrower term to 'Education' or 'Literacy education'. Of course, this is a very simple example and possibly meant for a subject expert who is knowledgable enough to comprehend the actual meaning of the subject. But there could be hundreds of subjects, especially in micro documents dealing with disciplines such as, chemistry, biochemistry, genetics, etc., where the index might need to display various rounds of Entities and Actions in a single index entry with their respective Properties (mostly as Modifiers, Parts and Constituents), eventually leading into much more confusions. There might even be occasions when the subject expert may find it difficult to ascertain the context-dependency of such Lead Headings.

Similarly, in the case of type 4, there could be occasions when certain entries might carry redundant information in them. For example, let us consider the following name of subject 'In information science, telecommunication-based online retrieval of statistical data'. This would be represented as per DS of SIL as:

Information science 8 Data 8.6 Statistical data 8.1 Retrieval

8.1.6 Online retrieval 8.1.6 Telecommunication-based online retrieval

Selecting the last component term falling in each of the ECs would give rise to the following 'Short Context Heading':

Information science 8 Statistical data 8.1 Telecommunication-based online retrieval

The component terms selected to form the Short Context Heading are used to form Upper Link Specifiers. The sequence of component terms in the Lead Heading containing Upper Link Specifiers taken from left to right is the reverse of the sequence of component terms in the Short Context Heading. Accordingly, we would get the following index entries:
The last entry certainly brings in an amount of redundant information, which could be described as the information which the searcher already has, because the Context Heading merely repeats the same information conveyed by the Lead Heading, the difference being only in their format and sequence.

Normally, a POPSI (as well as DSIS) index is either 'bipartite' or 'unipartite' (Bhattacharyya and Chandran, 1983, p. DC9; Devadason, 1986a, p. 34). A bipartite POPSI index consists of an 'Organising Part' complemented by an 'Associative Part'. But in a unipartite POPSI index, the Organising Part and the Associative Part are merged together. In the Organising Part of a bipartite index, each entry is under its modulated name of the subject (alternatively, under the Full Context Heading as 'feature heading'); and all the entries are arranged predominantly according to their alphabetical makeup being governed by the 'indicators of deep structure' (or in other words, alphanumerically). This produces an 'organising effect' in the sequence of the main entries (bibliographic references). The entries for the Associative Part are made of secondary subject propositions (i.e., Lead Headings alone) arranged on a purely alphabetical basis. Each entry bearing a secondary subject proposition consists of two distinct parts:

1) the approach-proposition (Lead Term); and

2) the referred to object (in this instance, any true substitutes of
the modulated subject propositions such as, the entry numbers).

The approach-propositions are all sought subject propositions. An approach-proposition may consist of:

1) A single facet-term (similar to Uni component term Lead Heading) or
2) Multiple facet-terms (similar to Lead Heading with Upper Link Specifiers).

In such a bipartite index things may not be as cumbersome as would be the case with the unipartite ('Associative-cum-Organising') one, because it will be easier to make a visual scan of the entries in the Organising Part of the index. However, as indicated in section 6.3.5 below, an online search system might be of some help to reduce the tedium of manual searches. Even, there may not be any further need to have a separate Associative Part (or 'Associative Grouping' feature) in the index. However, such remarks are still hypothetical by nature. Much research is needed in this direction to prove their validity. There is also the overriding need to prove that such an operation will be cost-effective, because apart from the likely increase in indexing costs, the use of indicators may prove costly in terms of search query formulation and actual search processing (Lancaster, 1972, p. 219).

Most of the studies on PRECIS have used the standard form of the index. The only other form of PRECIS index used is a one-stage index (i.e., without cross-references). Such a PRECIS index (on catalogue cards) comprised one of the test indexes used during the WUSCS experiment (Hunt et al, 1976/77). In comparison with other types of index, viz., LCSH and KWOC, as well as with other variations of PRECIS index, viz., with cross-references and in book form, it performed equally well. Similarly, the possibility of a one-stage PRECIS index for the British Education Index was evaluated as part of the wider survey of PRECIS user reaction in the UK (Peters, 1981).
6.3.4 Cross Reference Structure [3]

In DSIS, apart from cross reference entries to control naturally occurring synonyms, quasi-synonyms, etc., See cross reference entries are required to control synonyms artificially created by factoring composite terms, e.g.

Phthisis See Medicine (D) + Lung (E) + Tuberculosis (P).

[where D = Discipline, E = Entity and P = Property.]

In majority of indexing systems using cross reference entries (both See and See also types) the search process involves a two-step operation. First, the searcher is directed from the 'referred from' heading to the 'referred to' heading. For example,

Preserved Context Index System See PRECIS
POPSI See also DSIS

And the second involves a search for the relevant entries under the 'referred to' heading. As PRECIS does not recommend semantic factoring, there is no question of having such cross reference entries directing the user to look under the factored constituents of a composite term. In DSIS, this search process may become very complex for the average searcher, which may increase to more than two steps, especially in a manual searching system (e.g., in a printed index such as, the British National Bibliography subject index). If the decision is to produce an index with 'Uni component term Lead Heading with either Full Context Heading or Short Context Heading' (Devadason, 1986a, pp. 39-40: Exhibits 4 and 5), then the searcher will be first directed to look under the factored constituents of the Composite Term. But in the next step of search the searcher will have to take a decision, whether to look under the Discipline term 'Medicine' or the Entity term 'Lung' or the Property term 'Tuberculosis'. As the Discipline terms are
not usually selected to form Lead, the searcher can restrict his choice between either 'Lung' or 'Tuberculosis'. But again the choice will be immaterial. Since, there are no 'Upper Link Specifiers' (Qualifiers) to the Lead terms, the searcher will have to scan through the Context Heading Sections under either Lead term and look for a combination of 'Medicine + Lung + Tuberculosis'. Even then, there is little guarantee that these terms will be consecutive to each other in the Context Headings. Besides those intervening superordinate terms to each EC manifestation in the name of subject, there is always the possibility that, the Discipline and Entity terms may be interpolated by some other Entity terms (mostly as Modifiers) representing 'Systems' (such as, 'Ayurveda', 'Homeopathy', etc) and/or 'Specials' [2] (such as, 'Child', 'Female', 'Tropical', etc). Similarly, the Entity term 'Lung' and Property term 'Tuberculosis' may be interpolated by some terms denoting the former's Parts such as, 'Upper lobe', 'Lower lobe', etc. For example, consider the following Context Headings:

Medicine 9.6 Ayurvedic medicine 8 Respiratory system 8.4 Lung
8.2 Disease 8.2.6 Tuberculosis 8.2.1 Treatment 8.1.6 X-ray
treatment

Medicine 9.6 Child medicine 8 Respiratory system 8.4 Lung
8.4 Upper lobe 8.2 Disease 8.2.6 Tuberculosis 8.2.1
Treatment 8.1.6 Surgical treatment

Medicine 9.6 Child medicine 9.6 Ayurvedic child medicine
8 Respiratory system 8.4 Lung 8.4 Lower lobe 8.2 Disease
8.2.6 Tuberculosis 8.2.1 Treatment

Of course, these are very simple examples. In an index to a large and micro
document-based database, where there will be a large number of such Context
Headings under each Lead Heading, this search procedure is bound to be
difficult and tiresome. As a time-saving alternative to the above
procedures, the searcher can match the entry numbers bearing the search
terms 'Lung' and 'Tuberculosis' and select those entry numbers only which are common to both as might be done in postcoordinate indexing systems. But there will always remain the risk of 'false combinations'. For example, in a search strategy involving the above two terms, the following entries will also be retrieved:

**Lung**

- Medicine 8 Respiratory system 8.4 Lung 8.2 Disease 8.2.6 Bacterial infection 8.2.1 Diagnosis
- Medicine 8 Respiratory system 8.4 Lung 8.2 Disease 8.2.6 Blood coagulation 8.2.1 Prevention

**Tuberculosis**

- Medicine 8 Genito-urinary system 8.4 Kidney 8.2 Disease 8.2.6 Tuberculosis 8.2.1 Treatment 8.1.6 Surgical treatment
- Medicine 8 Nervous system 8.4 Brain 8.2 Disease 8.2.6 Tuberculosis 8.2.1 Diagnosis

However, only after the proper order is established, can the searcher start looking into other aspects of the search query related to the subject 'Phthisis'. In the case of an index with 'Lead Heading with Upper Link Specifiers and Short Context Heading' (Devadason, 1986a, p. 32), the searcher will find all the 'referred to' constituents of the factored Composite Term, side by side, in the Lead Heading Section of the entries, but in the inverted sequence of the terms as suggested in the 'referred to' section of the CR entry. For example,

**Tuberculosis, Lung, Medicine**

- Medicine 8 Lung 8.2 Tuberculosis 8.2.1 Prevention
- Medicine 8 Lung 8.2 Tuberculosis 8.2.1 Surgical treatment
- Medicine 8 Lung 8.2 Tuberculosis 8.2.1 X-ray treatment

On the other hand, if the decision is to produce an index with 'Lead Heading
with Upper Link Specifiers and Full Context Heading' (as we have decided for the purpose of this comparative study) (Devadason, 1986a, p. 32), the searcher will find all the 'referred to' constituents of the factored Composite Term, side by side, in the Lead Heading Section of the entries, but in the inverted sequence of the terms as suggested in the 'referred to' section of the cross reference entry. For example,

Tuberculosis, Lung, Medicine

Medicine 8 Human body 8.4 Respiratory system 8.4 Lung 8.2 Diseases
8.2.6 Tuberculosis 8.2.1 Prevention
Medicine 8 Human body 8.4 Respiratory system 8.4 Lung 8.2 Diseases
8.2.6 Tuberculosis 8.2.1 Treatment 8.1.6 Surgical treatment
Medicine 8 Human body 8.4 Respiratory system 8.4 Lung 8.2 Diseases
8.2.6 Tuberculosis 8.2.1 Treatment 8.1.6 X-ray treatment

The searcher needs to be aware of such formulations beforehand. There is also the need to explain all the components of an index entry including codes. The above cross reference entry structure is suitable for the Alphabetic Index Part to the Systematic Part of the Classaurus (Devadason, 1985b, p. 20), in which terms are arranged into separate hierarchic schedules of the elementary categories: Discipline (D), Entity (E), Property (P), and Action (A), together with their respective Species/Types, Parts and Special Modifiers. The above entry structure helps users of a Classaurus to find their way into the respective schedules of the elementary categories to which the constituents of a Composite Term belong. But it appears that this will be a considerable burden on the average searcher/user of an alphabetic subject index, to be aware of such cumbersome and importunate search procedures (Biswas and Smith, 1988b). Obviously, this brings in the question of the skill of the searcher, one of the external variables of a behavioural kind, which affects the search process (Svenonius, 1986, p. 331). Even, a
highly syntactical and rigorously controlled indexing language, such as DSIS, does not have complete control over it.

6.3.5 Display of Index Entries

In the display of the subject statement in PRECIS, once the string is formed, the output does not indicate how words in the subject statement are syntactically related. In other words, the operators and codes are deleted from the index entries. Instead, the responsibility of showing syntactical relations are delegated to the word order and certain connectives. But, in DSIS the relational indicators (excluding processing codes) are retained. Evidently, the logic for their retention is that they would show the syntagmatic (and paradigmatic) relationship of elements in a subject-proposition in a better way. The inclusion of relational symbols in indexing languages was advocated by Mineur (1973) also. But, certainly, "the use of relational symbols would make the output index very ugly and it is also doubtful how far this artificial syntax would be understood and utilised by the general users of the index" (Rajan and Guha, 1979, p. 380) as well as increasing its bulk. In the context headings (display in PRECIS) formed by DSIS, there are no special symbols other than Indo-Arabic numerals and the dot to indicate categories and their role, and an alphanumeric arrangement is an improvement over the pure alphabetical order, but it will be a bit preemptive to assume that the users' reactions will be favourable as well. In this respect, the structured display in a PRECIS index entry, using different print styles and few connectives (such as prepositions, conjunctions, etc., and punctuations, e.g., period, comma, colon, etc.) is more conducive and attractive to the user. From an indexer's point of view (as the experience of this author suggests), a PRECIS input string is much easier to construct, write and keyboard than a DSIS index. However, in an online search system the user can escape the mental drudgery of the manual
search and might be able improve the efficiency of the search process by using natural language words in conjunction with numbers representing 'indicators of deep structure' (instead of standard AND, OR, NOT Boolean operators). Devadason (1985a, p. 93) has informed us that investigation in this direction is ongoing. Similarly, a major research project into the use of PRECIS data for enhanced subject searching in static and especially online catalogues is currently in progress at Middlesex Polytechnic, UK (Congreve, 1986). The British Library has recently announced its commencing of a project to develop a new online interactive subject authority control system using the PRECIS system (British Library Bibliographic Services Newsletter, 1988, p. 8). Due to its online environment, the new system can do without the complex PRECIS manipulation coding, and its omission will make the system simpler and cost-effective. The software is fairly KWOC-type and capable of producing indexes comparable to those using PRECIS software.

Notes

1 The Non-hierarchical Associative Relationship "is a relationship in which terms are not equivalent and are not hierarchically related. The relationship includes among others, entities and their processes and properties, operations and their agents or instruments, actions and the product of the actions, the whole-part relationship other than the hierarchical whole-part, and many others" (Aitchison, 1986, pp. 164-165).

2 A Systems Basic Subject (or Systems) is "A Main subject expounded according to a specific System (of thought)". Whereas, a Specials Basic Subject (or Specials) is "A Main subject whose exposition is restricted to the special features of the entity concerned while within a specific environment or restricted in some other special manner" (Ranganathan, 1968, p. 100).

3 In this section the discussion on generation of cross references and
other related aspects are purposefully excluded, as they form part of the next chapter.


Devadason, F. J. (1986b). Personal communication.


7.1 Alphabetical vs. Classified Structure of a Controlled Vocabulary

It is a prerequisite of most controlled vocabularies that they must organize and display terms both alphabetically (to allow direct entry) and systematically (e.g., hierarchically) in a way helpful to both the indexer and the searcher (Lancaster, 1977). The major arrangement of vocabularies of more conventional kind, such as classification schemes, is hierarchical, but such schemes must also have an alphabetic index to allow access to the hierarchical schedules. For controlled natural language vocabularies such as thesauri, however, the main arrangement is alphabetical and the hierarchical structure is displayed explicitly within this (Jones, 1987, p. 118). The former constitute vocabularies of an a-priori kind, where the verbal representations for necessary concepts and concept relations in a document to be filed are expressed in one and only one permissible mode of expression before they are entered into the search file, and consequently their predictability is assured. Whereas, the latter could be used as either a-priori or a-posteriori vocabularies, since, these vocabularies can be established and updated either before or after the corresponding documents are filed.

There are a number of significant advantages of thesauri as compared with hierarchical classification schemes. Many classification schemes provide only limited hospitality for nascent concepts and concept relations that are continuously being created by research and development in the subject field concerned. Whatever effort is made to cater for such provisions, such an a-priori vocabulary is bound to be outdated as soon as it has been constructed. This problem may be further compounded by the
sacrifice of hierarchical accuracy to notational constraints and increasing separation of like topics accompanying increasingly specific division. On the other hand, a thesaurus provides an apparently unlimited hospitality for new concepts and their relations, and also is free from any notational rigidity. Besides those pure hierarchical relationships (e.g., genus-species, whole-part, broader subject-narrower subject), it offers the possibility of displaying polyhierarchical and cross-hierarchical relationships among terms (by means of RT or 'related term' references). It achieves high precision by distinguishing homographs (through the use of parenthetical qualifiers or by numbering (Soergel, 1974, p. 301)), and by providing definitions or scope notes for some of the terms. Due to these favourable features of a thesaurus, it has often been rated superior to bibliographic classification schemes (which lack them), especially with respect to its efficacy in tackling future changes and its ease of implementation in an information storage and retrieval system.

However, as a vocabulary using natural language terms, the thesaurus inherits several lacunae that necessarily belong to the natural language used for the purpose. Among these are the ambiguity of many terms, their instability in meaning over time, lack of common strings of characters (word stem, syllable, etc.) among closely related concepts to facilitate a generic search (as is possible in the case of systematic notations), simulation of concepts that are not implied, etc. These inherent lacunae of a natural language vocabulary, such as thesaurus, have been further complicated by documentalists in their over-enthusiastic attempts to include all newly emerging terms and their relations on a purely physical basis into it. As a result, the thesaurus soon exceeds the limits of its operancy and increasingly fails to serve the purpose of an efficient device for reliable terminological control at the indexing and retrieval stage. According to Fugmann (1974, p. 78), this pernicious and fatal process "can effectively be
counteracted by recalling to mind principles typical of analytico-synthetic classification and by employing them in a balanced and proper combination together with the thesaurus approach". It seems possible that, by combining their respective advantages, their disadvantages would cancel out, resulting in both high precision and high recall.

7.2 Facet Analysis and Vocabulary Control

As has been suggested by Fugmann above, the most typical principle of analytico-synthetic classification that comes to one's mind, is the principle of facet analysis, one of the most significant contributions of Dr S. R. Ranganathan. The procedure was first enumerated by Ranganathan (1944) and was subsequently incorporated in the skeletal form of five Fundamental Categories (viz., Personality, Matter, Energy, Space, and Time) in the fourth edition of his Colon Classification published in 1952. Since then, the technique of facet analysis has paved the way for the development of numerous faceted classification schemes and new indexing languages in India and elsewhere. Even, there is a tendency to adopt some level of 'faceting' in most of the traditionally enumerative classification schemes such as, the Library of Congress Classification, and explicitly in Dewey Decimal Classification.

Facet analysis involves sorting candidate terms in a vocabulary into homogeneous, mutually exclusive facets (categories) on the basis of a single characteristic of division, where a 'facet' has been taken as "A generic term used to denote any component -- be it a basic subject or an isolate -- of a Compound Subject, and also its respective ranked forms, terms and numbers" (Ranganathan, 1967, p. 88). Although classificationists used this procedure in the construction of classification schemes, it could be equally useful in designing and structuring of any type of controlled vocabulary including thesaurus. It can help in determining the elementary categories of
terms, thus building up the major hierarchies of the vocabulary, and in
displaying the most useful relations among terms across hierarchies (by
means of See also or RT references). Facet analysis also can help us to
understand the true meaning of terms, sometimes by factoring (decomposing,
breaking down) certain terms into their component or elemental concepts and
thus to understand their correct relationships with other terms. For
example, 'computer' indicates an 'instrument' for 'processing of 'data'.

This suggests that, it is also possible to organize thesauri in
completely different ways (for reviews see Lancaster, 1972, pp.43-45, 66-69;
Aitchison, 1972, pp. 72-82; Devadason, 1986c, pp. 136-137). The earliest use
of facet analysis to group descriptors of a controlled vocabulary can be
found in a description of a uniterm system by Waddington (1958). He used
Ranganathan's Colon Classification for this purpose. Facet analysis was used
extensively in the construction of American Petroleum Institute's (1966)
Subject Authority List. A rudimentary form of facet analysis is evidenced in
the construction of a thesaurus of buildings (Rostron, 1968). McClelland and
Mapleson (1966) examined the hypothesis that a classified thesaurus with
specific/generic relationships is of value in indexing anesthesiology
literature and found it valid. Another systematic approach to the
construction of a classed thesaurus using a form of facet analysis, deriving
the facets from relationships displayed in scientific glossaries, was used
by London (1965, 1966). The hierarchical relationships of the Exploration
and Production Thesaurus (University of Tulsa, 1968) were also developed
through the use of facet analysis. Fundamental facet groups have been used
for a long time as a tool in the process of thesaurus compilation. Dym
(1967) used fundamental facet groups, such as 'materials', 'equipment',
'supplies', 'process and materials' and 'property, characteristic or
condition' in the production of a thesaurus for paint technology. Similar
broad groups or facets, known as 'themes', have been used by Jean Viet in
the layout of numerous thesauri for international organizations, viz., the *Thesaurus for Information Processing in Sociology*, the *International Thesaurus of Cultural Development*, the *Thesaurus -- Mass Communication*, the *EUDISED Multilingual Thesaurus for Information Processing in Education* (1973 edition), the UN-OECD *Macrothesaurus*, and early editions of the *ILO Thesaurus* (Aitchison and Gilchrist, 1987, p. 83).

Research has shown that, it is possible to develop a thesaurus having two complementary and entirely coequal parts, one a classification scheme showing formal hierarchical relationships and the second, an alphabetical thesaurus, showing the interhierarchical relationships that might be useful in indexing and retrieval. Barhydt et al's (1968) *Information Retrieval Thesaurus of Education Terms* is a detailed and explicit example of the use of facet analysis. The major example of course is the *Thesaurofacet* of Aitchison et al (1969), which incorporates a carefully constructed faceted classification scheme along with a complementary alphabetical thesaurus arrangement, and is thus able to eliminate their specific drawbacks, but at the same time preserve the full advantages of both. The classification part displays the relations that are best displayed through formal facet analysis while the thesaurus part shows the needed interfacet relations. Specific faceted classification schemes such as the *Bliss Bibliographic Classification* (BC2) (Mills and Broughton, 1977- ), have also been used as a source of terms and structure by several thesauri. Most of these were compiled in the United Kingdom. "Some used BC2 as one source among many, as was the case of the British Standard Institution's *ROOT* thesaurus, the *UNESCO* thesaurus, *Thesaurus on youth*, and the *Community information classification and thesaurus*. However, two thesauri, the DHSS-DATA thesaurus and the *ECOT* thesaurus are derived largely from BC2, and include classified displays based on the BC2 schedules" (Aitchison, 1986, p. 171). Croghan's (1970) *A Thesaurus-classification for Non-book Media* and the U. K. Depart-
ment of Environment's Construction Industry Thesaurus (Gilchrist, 1972) belong to the same category of controlled vocabularies. The most recent addition to this category is a thesaurus on equipment for disabled people, developed by the Disabled Living Foundation, UK (Mandelstam, 1988). In a similar vein, Wall (1972, 1973) described a method of using traditional classification schemes such as DDC and UDC, to generate a hybrid 'thesaurus-classification' by a form of 'learning process' in the course of regular indexing. In a later update, he (Wall, 1980) mentioned some similar attempts, prominent among which are the works of Alber (1972), Batteké et al (1974), Field (1974), Schreiber et al (1975), Gorol et al (1975), Ghose and Dhawle (1979), and Hindson (1979). Most of these works were on the desirability of thesaurus-UDC combinations in automated systems.

In the United States, Batty (1981) evidenced a renewed interest in facet analysis forty years after its first full exposition by Ranganathan. "The tree structure of MeSH is a simple and well known example, now followed in other index languages, for example TEST and ERIC (until the most recent edition). ...BIOSIS is now using a faceted approach in automatic support indexing. A major effort is now in hand for the U. S. Department of Energy to create a data resources directory, indexing down to the data element level, using an eight-faceted scheme to describe elements". Batty himself developed a model system for a U. S. House of Representatives Subcommittee, using a simple controlled vocabulary contained in five facets, which showed notable increases in the speed and efficiency of vocabulary development, indexer training, and searching capabilities over conventional indexing. In Europe, facet techniques have been used by Bauer (1967) in chemical thesaurus construction, and by Marlot and Moureau (1970) for a petroleum industry thesaurus. Fugmann (1974) showed that the inherent features of both the thesaurus and the classification approach could effectively complement each other, and used this idea as a theoretical foundation for his IDC
The Polish Thesaurus of Hospital Science was created by Jachowicz (1979) by using the already existing faceted classification scheme for the same subject field.

In India, subject headings structured according to postulates and principles for facet analysis have been used for computer aided construction of thesaurus (Devadason and Balasubramanian, 1978, 1981). Based on the analysis of modulated subject headings according to POPSI (Postulate-based Permuted Subject Indexing) system, a system for generating an information retrieval thesaurus using computer has been designed (Devadason, 1983). A new vocabulary control and indexing tool called 'Classaurus' has been developed based on the analysis of subjects according to the POPSI system (Bhattacharyya, 1981a, 1982a). It is a faceted, hierarchic scheme of terms with conventional vocabulary control features, designed on the basis of the General Theory of Subject Indexing Language (Bhattacharyya, 1979b, 1980), of which POPSI is a specific version. A methodology for the design of a species of Classaurus called 'alphabetic Classaurus', using computers has also been developed (Devadason and Kothanda Ramanujam, 1982). The problem of obsolescence associated with such a controlled vocabulary has been overcome by devising a system for online construction and updating of the alphabetic Classaurus (Devadason, 1985).

In many respects, the BNB/PRECIS thesaurus is the precursor of BSI ROOT thesaurus (British Standards Institution, 1985). In this sense, it can be categorized with the family of thesauri typified by the latter thesaurus. The systematic section is organized as a number of hierarchies, each term in such hierarchies being identified by a unique number (i.e., RIN). In the alphabetical display section, each term is again organized along with its immediate broader terms (BTs), narrower terms (NTs) and associated terms (RTs), derived from the machine-readable codes (such as, $o, $m, $n, etc.)
included in the RIN-file. The following sections try to compare the features generated from the respective structures of vocabulary control tools used by PRECIS and DSIS, along with the implications of some of the decisions taken in the construction of entries, discussed in the earlier chapter.

7.3 Vocabulary Control in PRECIS and DSIS

7.3.1 Term Structure

A vocabulary, whether natural language-based or controlled, alphabetically arranged or classed, is composed of 'terms'. In language, terms are used to designate concepts. So it is in a controlled vocabulary such as, thesaurus, subject heading lists, etc. A 'term' is the verbal representation of a 'concept'. Concepts -- unitary and compound -- may be represented by either single-word terms (e.g., 'Personnel', 'Hypertension' etc.) or multi-word terms (e.g., 'Skilled personnel', 'Labour productivity', etc.) (Seetharama, 1976, p. 13). Although, there is no one-to-one correspondence between concepts and terms in natural language, in a controlled vocabulary these are used interchangeably, since it applies terminological control through its homonym-synonym structure (Soergel, 1974, p. 20). As far as the problem of terminology is concerned, single-word terms representing unitary concepts present no problem at all in pre-coordinate indexing systems such as PRECIS, DSIS, etc. However, things are not that clear once we enter the domain of multi-word terms representing both unitary and compound concepts. Single-word terms representing compound concepts also have their own share of problems. In the previous chapter, we have discussed the problems related to the treatment of multi-word terms representing unitary concepts (prepositional phrases) in both the systems. The present section in this chapter will concentrate mainly on the the other major category of terms, viz., compound terms, both single-word and multi-word. Both ISO 2788 (International Organization for Standardization, 1974) and BS 5723 (British
Standards Institution, 1979) recommend that, as a general rule, terms should represent simple or unitary concepts as far as possible, and compound terms should be factored (decomposed) into simpler elements except when this is likely to affect the user's understanding. Complex subjects should be expressed by combinations of separate terms, which may function as components of pre-coordinated index entries. However, it was stipulated that this should be regarded as a recommendation rather than a rule. Factors such as (a) literary warrant, (b) indexing in a special field, and (c) need to regulate the number of postings per document, or the number of terms in the indexing vocabulary, may lead to the need for modifications. According to Jones (1981, p. 60), the first and third have been overtaken, or are about to be overtaken, by the move away from manual to computerized systems. The second may be true, for example, the highly synthetic vocabulary of chemistry may assimilate extensive factoring, but for social science vocabularies it may be appropriate. In the face of such criticisms, the later editions of both the standards (ISO 2788-1986 and BS 5723:1987) withdrew these factors as possible reasons of deviation from the rule, but still mentioned that the requirements of the individual indexing situations should be taken into account while applying the factoring rules. BS 5723 described two techniques for factoring of compound terms in indexing:

(a) syntactic factoring, described as a technique applied to compound terms, i.e. "terms which are amenable to morphological analysis into separate components, each of which can be accepted as an indexing term in its own right", e.g.,

Book indexing = Books + Indexing
Money supply = Money + Supply

(b) semantic factoring, described as a technique applied to "A term which expresses a complex notion is re-expressed in the form of simpler or
definitional elements, each of which can also occur in other combinations to represent a range of different concepts", e.g.,

*Pneumonia = Lung + Bacterial infection*

*Thermometers = Temperature + Measurement + Instruments*

ISO 2788 also described these techniques in a similar fashion. The latter technique is not recommended for pre-coordinate indexing purposes in either of these standards and it is indicated that it is "recognized that semantic factoring leads to a loss of precision in retrieval". Besides, anything like this moves further away from natural language. The factoring of compound terms in PRECIS conforms to the above rules and implications of these rules are depicted in the PRECIS thesaurus. Now taking these recommendations put forward by standards as a sort of background, we shall proceed to the discussions of the term structure recommended for DSIS and its implications for the Classaurus in general and 'Alphabetic Classaurus' in particular.

7.3.1.1 Logicality of the DSIS' Term Structure

We have seen that DSIS distinguishes three different types of term for the purpose of indexing, viz.:

1) Compound Term;
2) Complex Term; and
3) Composite Term,

"as the basic building blocks of subject headings" (Devadason, 1986a, p. 33). The same constitute the basic building blocks of the Classaurus. However, serious doubts could be raised about the nomenclature and logicality of such division (grouping). As "Classification (indexing) is always specific purpose oriented" (Bhattacharyya, 1981, p. 96), it might be useful if we look into the literature of classification (indexing) for an understanding of the notions of the above-mentioned concepts, rather than
applying their meaning in common usage. There are no made-to-measure definitions for these concepts in Prolegomena (Ranganathan, 1967), the source-book of modern classificatory terminology. However, it is possible to derive them from related definitions. The nearest concepts one could find are the Compound Class Term, Complex Class Term and Derived Composite Term, which respectively parallel the ideas of the Compound Term, Complex Term and Composite Term in DSIS. According to Ranganathan (1967, p. 86-87), a Compound Class Term is a "Term denoting a Compound Class; it is the name of the Compound Subject" [1] and "may consist of a single term consisting of a word or word-group, or it may consist of a succession of blocks of a basic term and isolate terms". For example,

Paediatrics (single term single word)
Information retrieval (single term multiple word)
Geology of sedimentary rocks in Eastern India (succession of blocks of a basic term (subject) and isolate terms).

Whereas, a Complex Class Term is a "Term denoting a Complex Class; it is the name of the Complex Subject" [2] and "usually consists of the names of the component class term connected by some words as in 'Physics compared with Chemistry'". But "It may occasionally be a single word or word group as in 'Geo-politics' standing for 'Political science influenced by Geography'". So it seems that structurally both the Compound Class Term [3] and the Complex Class Term can often be similar. Thus one can say that, Ranganathan's mode of grouping was conceptual rather than structural. Now, by introducing the same logic of grouping at the 'isolate' [4] level to both the concepts of Compound Class Term and Complex Class Term, we can deduce the construct of the Compound Isolate Term [5] and Complex Isolate Term [6]. Reflections of Ranganathan's above mode of grouping at the 'isolate' level of classification (indexing) were also evident in the following illustrative
schedule of Compound Isolates. It is reproduced from chapter 'DE Language Isolates' of the 7th edition of Colon Classification (1971) (Ranganathan, 1970, p. 24-25) [7]:

111-k Technical jargon of English Division by (SD) (illustrative) 111-A Stages of English language
111-k(B) for Mathematics 111-D Old English
111-k(C6) for Electricity 111-E Middle English
111-k(Z) for Law 111-J Modern English

111-J-d185 Modern Yorkshire

English

Devadason (1986a, p. 4) has, in fact, coined the term 'Complex Term' on the basis of Ranganathan's 'Complex Class Term' and 'Complex Isolate Term'. So one could safely presume that as a natural logical progression, the grouping of modified term into Compound Term and Complex Term would follow the same criterion. But we see that the 'structure' of the 'Modified Term', rather than its 'conceptual base' (semantic significance), has been selected by Devadason (p. 3) for the grouping of Modifiers.

The British Standard (BS 5723, p. 9) on vocabulary control and thesaurus construction, does not differentiate between Compound Term and Complex Term, and labels them together as Compound Term (of course, excluding complex subjects formed by phase relations). According to this standard (p. 5; as we have seen earlier), noun phrases belonging to the category of Compound Term occur in two forms:

(a) *adjectival phrases:*

Skilled personnel, Income taxes, Lending libraries, etc.

(b) *prepositional phrases:*

Cataloguing-in-publication, Management by objectives, etc.

In English language grammar (Quirk et al, p. 1239) these are characterised
as the 'premodification' and 'postmodification' of the 'noun phrase'. Their conceptual proximity could be assessed from the fact that in many cases, premodifying components (as 'Skilled' in 'Skilled personnel') in the compound terms correspond to postmodification with prepositional phrases and vice versa, e.g.,

Skilled personnel  ~  Personnel with skills
Income taxes  ~  Taxes on income
Lending libraries  ~  Libraries for lending

[Note: "~" indicates systematic correspondence between structures.]

"It is probable that most compound words evolve from phrases, into loose compounds, then into compounds where the words are separated by hyphens, and finally into bound words... More than one form may exist concurrently within a literature" (Jones, 1981, p. 54). In modern linguistics this has been labelled as the 'surface structure' of a particular natural language (in this case English) which uses adjuncts and conventions to convey the basic inter-concept relationships, alternatively known as the 'deep structure' level of all natural languages. "Surface adjuncts and conventions (but not deep structure relations) vary from language to language, which explains why different languages need their own prescriptive grammars and their own rules for recognising well-formed sentences" (Austin, 1984, p. 299). For example, some languages express relationships through prepositions (as in German), whereas others use postpositions (as in Finnish); some inflect their nouns (as in Slavonic languages); others, such as Chinese, depend mostly on word order. Many languages use combinations of these different devices -- German, for example, possesses a rich repertory of prepositions and also inflects its nouns. In a similar manner, Devadason (1986a, p. 3) has also noted that, "those modifiers forming a Compound Term in one natural language (as in Hindi) may require the insertion of
auxiliary/function words and form a Complex Term in another natural language (as in Danish) and vice versa. Hence, the above grouping of Modifiers depends on the natural language concerned (exactly, its surface structure) from which the index terms are selected. But contrary to this logic, Devadason has included this feature as two of the subdivisions to the list of the Elementary Categories (derived on the basis of the semantic significance of the substantives) and prescribed .5 and .6 as indicators of the Complex Term and Compound Term, respectively. It could be said that, this feature would have been handled more appropriately by using some secondary codes (such as the Process Codes), rather than assigning two separate primary operators (indicators) for each (Biswas and Smith, 1988b). Because, deep down in their nature both signify the same conceptual (semantic) category. Jones (1981, p. 67), who studied the problems associated with the use of compound words in thesauri, concluded that for the purpose of factoring (or determining the nature) "criteria based on semantics appear to be more satisfactory than those based on syntax". In all language systems, natural or artificial, the syntactical and semantic structures overlap to a large extent (Austin, 1984, p. 70). Consequently, there is a high degree of relativity in the semantic-syntactic distinction within an indexing language. According to Gardin (1973, p. 145), "this dichotomy can be shown to have no justification other than practical. A given relationship between two or more concepts can be expressed either analytically (i.e., semantically)... or syntactically... the choice... depends on practical circumstances that are of the same nature in ML (meta language, such as an indexing language) as they are in NL (natural language): to be very brief, the more straightforward relations, those which 'go without saying' but which are nevertheless frequently put to use in a given universe of discourse tend to be transferred to the semantic structure, whereas the more unpredictable relations, bringing forth new knowledge, find their expression in the syntactical structure". But,
fortunately or unfortunately, there is a mix up in the use of 'indicators' in DSIS. Indicators such as, .4 (Part), .5 and .6 (both denoting Species/Type) represent the semantic structure in the name of a subject. Whereas, the remaining represent the syntactic structure such as, Discipline-Entity, Entity-Property, Entity-Action, Entity-Place and so on. In this context, the case of .5 indicator seems to be a marginal one, involving an overlap between the two structures. It is deemed to represent a Species/Type of the modifyee, but in reality, it is also being used to indicate the relationship between two separate components in the name of subjects. It is hard to agree that, concepts such as, 'Water repellancy (agent used) Silicone' is a Species/Type of 'Water repellancy'. The term 'Water repellancy' denotes a Property/Process and 'Silicone' is an agent to achieve that effect. Their relation could be best described as one of the 'non-hierarchical associative' type (represented by RT or 'related term' in conventional thesaurus) (Aitchison and Gilchrist, 1987, p. 46). The second part of the phrase, 'Silicone', should have been represented by a new Elementary Category (EC) and indicated by some syntactic device, may be a new indicator. This is quite different from the concept of 'prepositional noun phrase' creating Species/Type, discussed above.

In DSIS complex subjects formed using auxiliary/function words are also categorised along with the multiple-word terms using prepositions in them (Devadason, 1986a, p. 4). The implication of this rule is that, there will be Complex Terms such as, 'Education (effects of) Technological change', 'Adult education (role of) Mass media', etc., which are hardly conceivable as single terms, even by a specialized user. Apart from the terminological constraint faced by DSIS in its attempt to overgeneralize the use of Modifier of Kind 1 by including Complex Subjects formed by phase relations (Biswas and Smith, 1988a, p. 7; also Section 6.1.3, Chap. 6), there is a logical fallacy in Devadason's argument. According to Hutchins, (1975, pp.
104-105), phase relations, such as 'influence', 'comparison', 'bias', etc., are not to be regarded as internal relations (semantic relation) of descriptions of subject content. Rather, they are of external type, concerned with some pragmatic aspects of an indexing language, such as to illustrate a particular perspective through which a subject is studied, to indicate the kind of audience for which the document is written, etc. These clearly stand outside the representation of semantic content of the subject. Similar views were also held by Coates (1973, p. 393) for whom "Of the four Ranganathian Phase Relationships, three, namely Bias, Comparison, and Tool phase were eventually elucidated as connections between concepts at the level of the author's treatment rather than inherent logical relations. Only Influence Phase remained as a true syntactic relation". Incongruously with this logic, Devadason has drawn the inference that Complex Subjects formed by phase relations are generally narrower than the subject represented by the first phase, hence, they also could be considered as Species/Type, same as the prepositional phrases. In brief, the relation between the first phase (e.g., 'Education') and its so-called Species/Type (e.g., 'Education (effects of) Technological change') is one of semantic nature, which it isn't. Thus, the above argument raises the fundamental question of adequacy of the number of ECs used in DSIS to represent the class of concepts occurring in the name of subjects.

Last but not least important, is the idea of the Composite Term. We have seen that, if a component term in a name of subject represents manifestations of more than one Elementary Category, then it is a Composite Category Term or simply Composite Term (Devadason, 1986a, pp. 6-7). Such Composite Terms are to be broken down (factored, decomposed) into their fundamental constituent terms or elemental concepts [8] and each one of them identified as belonging to one or the other of the ECs. According to Ranganathan (1967, p. 87), when a Compound Class Term consists of a single
term comprising a word or word-group, it is called a 'Derived Composite Term' [9]. "A derived composite term really represents a compound focus. It should be broken down into its component simple foci" (italics mine) (Ranganathan, 1968, p. 130). BS 5723 (p. 10) also treated such Composite Terms as other types of the Compound Term, rather than designating them as an altogether separate class of term. Because, conceptually again they belong to the same level as that of the idea of Compound Term and Complex Term, discussed in the previous paragraphs. The standard recognized the need for factoring such terms following the same techniques as those mentioned above in Section 7.3.1.

7.3.1.2 Implications of Term Structure of DSIS vis-a-vis PRECIS

In this section we will take the above division of terms or concepts as granted and rather concentrate on the practical impact it will have on the resulting Classaurus. Side-by-side, comparisons will be carried out with the similar aspects of the PRECIS' vocabulary. For convenience, each category of terms will be considered in turn.

A. Compound Term/Adjectival Compounds

In DSIS, the Compound Term is created by the application of Modifier of Kind 2 to the manifestation of an EC, which usually creates a Species/Type of the modifyee (focus). This takes the form of an adjectival phrase only (not also the prepositional phrase as defined in ISO 2788). For example, a Compound Term like 'Fee-based online bibliographic information retrieval services' is a type of 'Information retrieval service'; the focal noun term 'Information retrieval services' is being modified by three successive modifying adjectives, viz., 'Fee-based', 'Online' and 'Bibliographic'. This Compound Term would be input as,
Information retrieval services $3 Bibliographic $3 Online $3 Fee-based

and rendered as

Information retrieval services. Bibliographic information retrieval services. Online bibliographic information retrieval services. Fee-based online bibliographic information retrieval services.

This gives rise to the following hierarchy

Information retrieval services
  .Bibliographic information retrieval services
    ..Online bibliographic information retrieval services
      ...Fee-based online bibliographic information retrieval services

and the resulting Classaurus entry (incorporating similar sub-hierarchies) will be

Information retrieval services
  .Bibliographic information retrieval services
    ..Online bibliographic information retrieval services
      ...Fee-based online bibliographic information retrieval services
        .Fee-based information retrieval services
          .Online information retrieval services
            ..Fee-based online information retrieval services
              ..User-friendly online information retrieval services

In a similar manner, the same term will be treated in PRECIS as either a multiple-level differencing term, as follows:

*(1) information retrieval services $21 bibliographic $21 online $21 fee-based

and would generate the following leads with the whole term as their display:
Information retrieval services
Bibliographic information retrieval services
Fee-based information retrieval services
Online information retrieval services

or, the following hierarchy of terms (as displayed in the alphabetical section of the thesaurus):

Information retrieval services
NT Bibliographic information retrieval services
   Fee-based information retrieval services
   Online information retrieval services

and would generate See also reference entries under the broader term relating all the narrower terms to it, as follows:

Information retrieval services
See also
   Bibliographic information retrieval services
   Fee-based information retrieval services
   Online information retrieval services

It is worth pointing out that, in the Classaurus hierarchy the term 'Fee-based online bibliographic information retrieval services' has not been repeated under either 'Fee-based information retrieval services' or 'Online information retrieval services'. As a result, when the indexer [10] looks through these two latter terms, the hierarchy will not lead down to 'Fee-based online bibliographic information retrieval services', a logical subdivision of the above two terms. Even if the search begins through the alphabetical index part, the result will be the same. Only when the search starts from the top superordinate term 'Information retrieval services',
will the indexer be able to locate it under 'Online bibliographic information retrieval services' through the intermediate step 'Bibliographic information retrieval services'. But this is sufficient to confuse. As Classaurus does not permit the inclusion of associatively related terms, there is no way to relate it with either 'Fee-based information retrieval services' or 'Online information retrieval services'. Due to alphabetic adjacency, it might be found in the alphabetic index part close to the entry 'Fee-based information retrieval services', but not to the entry 'Online information retrieval services'. It seems that, the root of this problem lies in the system's failure to recognise the order of words in English adjectival compounds (as shown in Section 6.1.1, Chap. 6) as well as the fact that a compound term representing two (or more) different principles of division should always be factored (Lancaster, 1986, pp. 56-57). Our conviction will be still stronger if we consider the following example:

```
1st level differences      -- focus
/    \                   
(Toshiba) -- (portable) -- (computers)

or

(Portable) -- (Toshiba) -- (computers)

Figure 7.1: Levels of Modification
```

The above adjectival phrases appear to be natural and represent the same concept, though structurally different from each other. But this latter factor will give rise to different and unrelated hierarchies in the Classaurus, because "DSIS does not worry about direct and indirect modifiers as it totally depends on the hierarchy and other relations as presented by the document being indexed" (italics mine) (Devadason, 1986b, pp. 4-5). This will subsequently undermine the retrieval capability of the vocabulary. For example,
If combined together, it will be

Computers
.Portable computers
..Toshiba portable computers
.Toshiba computers
..Portable Toshiba computers,

which is self explanatory. According to Ranganathan (1967, p. 429), this problem of order among quasi-isolate ideas forming a superimposed isolate idea can be resolved by using the Wall-Picture Principle which will amount to the sequence 'Quasi-Isolate Idea purpose. Quasi-Isolate Idea brand' [11], i.e., 'Portable Toshiba', but not 'Toshiba portable'. The new hierarchy now will be,

Computers
.Portable computers
..Toshiba computers
..Portable Toshiba computers

But again, when one searches for 'Portable computers' one will not reach 'Portable Toshiba computers', except that both may appear close to each other in the alphabetical index. But one cannot rely on chance factors such as 'alphabetic adjacency' for consistent retrieval purposes. Devadason (1986b, p. 3) further suggests that 'If by chance, the alphabetical adjacency technique does not work, then permuted (inverted) renderings of Compound Terms may have to be included in the alphabetical index to the
Classaurus". So where are we? We began with a set of rigorously defined categories derived on the basis of Deep Structure of Subject Indexing Languages (SILs), which itself forms part of the General Theory of SIL developed mainly through logical abstraction of the principles of classification. And, ended up by taking sanctuary in the simplest order of alphabets (what Metcalfe (1957) calls the 'world's most important precision tool'!) and also resorting to as mechanical a device as word permutation. But "the order of words in adjectival compounds is more than a matter of chance, and cannot be explained only in terms of speech conventions (or word orders as presented by documents). Instead, this order appears to be subject to rules which can themselves be related to an internal categorical system" (Austin, 1984, pp.309-310). Such a system of categories was established empirically by Austin to determine the preferred order of components in an adjectival compound in PRECIS (Preserved Context Index System). The need for a similar set of rules was felt during the construction of Classaurus (Biswas and Smith, 1989).

B. Complex Term/Complex Subject

For Classaurus, there are some serious practical implications of the above treatment (cf. Section 7.3.1.1) of terms having phase relationships and/or facet relationships among them as Modifiers of Kind 1, creating Complex Terms. Take, for example, the portion of the Discipline schedule of the proposed online alphabetic Classaurus of education terms, generated during our project, and as enumerated below:

Education
(effects of)
-Technological change
-Unemployment
Education
(related to)
-Racism
-Adult education
(in)
-Prisons
(using)
-Interactive videos
.Basic education
...Adult basic education
(role of)
-Mass media
-Voluntary organizations
.Continuing education
=Lifelong education
=Post-compulsory education
etc.

We have been informed that "A modifier can modify a manifestation of any one of the ECs (even Discipline), as well as a combination of two or more manifestations of two or more ECs" (Devadason, 1986a, p. 2-3). The above hierarchy of Discipline schedule of education shows the outcome of above proposition. It displays an array of Sub-disciplines generated by both Modifiers of Kind 1 and 2, respectively. There were altogether 30 Complex Terms (formed by Modifier of Kind 1) forming the hierarchies of the Discipline schedules of the subject fields of 'Education' and 'Information retrieval', majority of which, of course, occurred in 'Education'. There may be a general consensus over the propriety of adjectival compounds (e.g., Basic education, Adult basic education, Continuing education, etc.) forming Sub-disciplines of the focal noun 'Education', considered to be the main
Discipline (there were 143 occurrences of such adjectival compounds in the whole file, again majority of them in 'Education'). On the contrary, there can be a common reluctance among educationists to accept Complex Terms (e.g., 'Education (effects of) Technological change', 'Education (related to) Racism', etc.) also as the so-called Sub-disciplines (?) of 'Education'. The confusion created by the inclusion of what appear clearly to be examples of terms having facet relations (e.g., 'Adult education (using) Interactive videos') and having phase relations (e.g., 'Education (effects of) Technological change') in them, in the same class of terms along with those generated by prepositional phrases (e.g., 'Adult education (in) Prisons'), has in fact wiped out the initial advantages of strictly faceted hierarchic scheme of terms, such as Classaurus. The overall findings of this study suggests that these types of relation should be treated in a different way, may be through facet relation and/or inter-subject relation between terms (as in the version of POPSII described by Bhattacharyya (1981) in the DSIS index entries, and be listed as an altogether different category in the Classaurus. PRECIS does exactly the same things and its thesaurus simply categorizes them as the member of one of the basic class of concepts, viz., 'Entities' and 'Attributes'.

C. Composite Term/Factoring of Compounds

DSIS recommends that a Composite Term, representing manifestations of more than one EC, should be factored into its elemental concepts. PRECIS, following ISO 2788 or BS 5723, also recommends factoring of such concepts. However, there is a basic difference. PRECIS adheres to the recommendations of standards regarding 'semantic factoring', i.e., not to use them at all, because such artificial factoring leads to a loss of precision in retrieval. But, Devadason (1986a, p. 7) has failed to recognize the need for distinguishing semantic factoring and syntactic factoring, and it appears,
from the examples used by him, that he recommends both for the purpose of indexing and retrieval. The use of elemental descriptors for the representation of compound concepts gives better recall, but not high precision, because they are less specific. Recall too may be lost if the indexers and searchers use a different combination of terms to indicate the same concept, e.g.,

Concept: University library administration
Indexed: UNIVERSITIES and LIBRARY ADMINISTRATION
Searched: UNIVERSITY LIBRARIES and ADMINISTRATION

However, this recall failure can be avoided by the provision of an adequate entry vocabulary.

7.3.2 Representation of Basic Thesaural Relations

Thesaural relations, also known as a priori relations between concepts, are document independent and consist of three basic classes, viz.: (a) 'equivalence', (b) 'hierarchy', and (c) 'associative'.

7.3.2.1 Hierarchical Relations

Hierarchy is an organised list of terms based on the degree of subordination and superordination and includes genus-species, whole-part and broader subject-narrower subject relationships. For example,

<table>
<thead>
<tr>
<th>Plants</th>
<th>Great Britain</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trees</td>
<td>England</td>
<td>Biology</td>
</tr>
<tr>
<td>Conifers</td>
<td>Leicestershire</td>
<td>Botany</td>
</tr>
<tr>
<td>Firs</td>
<td>Loughborough</td>
<td>Plant pathology</td>
</tr>
</tbody>
</table>

The PRECIS thesaurus takes the responsibility of generating and displaying these relationships. In a PRECIS index these appear as single-step See also
cross references, e.g.

Great Britain

See also

England

Northern Ireland

Scotland

Wales

In DSIS (as in POPS1) the hierarchy is represented in the index entry itself, not through cross references. Each component in the name of subject is 'modulated' by augmenting it by interpolating and/or extrapolating as the case may be, the successive superordinates of each elementary category manifestation, by finding out "of which it is a species/type or part or constituent". For example, a subject such as 'Cataloguing of maps in university libraries' will be input as,

Library science 8 Libraries 8.6 Academic libraries 8.6 University libraries 8.4 Documents 8.6 Cartographic materials 8.6 Maps 8.1 Documentation 8.1.6 Cataloguing

As a result, DSIS has avoided the need for providing See also cross reference entries of the 'ascending' (from narrower to broader subjects) or 'descending' (from broader to narrower subjects) types in general, except in very special cases. Consider the context headings given under the lead term 'Basic education' in Appendix 4.1 in Chapter 4. All the full context headings having this term are grouped together, which automatically brings together all the context headings having narrower terms to 'Basic education' also, such as 'Adult basic education', 'Literacy education', etc. If one searches 'Basic education' he need not be directed to search also under narrower terms to it, using See also entries of the descending type like
PRECIS (Austin, 1984, p. 210). Similarly, the full context headings given under the lead term 'Adult basic education', also have the broader term 'Basic education'. If the searcher wants information on broader terms to the term he is looking at, he gets those terms from the full context heading under the term. Hence, there is no need for See also entries of ascending type. This approach results in a reduction in the number of times a searcher has to look up a new expression to continue a search, i.e., lesser number of two-step searches. The process of creating organizing classification effect in the verbal plane (i.e., by referencing) has been criticized as a case of "running from pillar to post" (Bhattacharyya, 1981, pp. 98-99). Computerized information retrieval from machine-readable databases using the same procedure may seem to be a solution to this problem, because of the speed and the least effort on the part of the searchers. But this solution is more apparent than real. On the other hand, the alternative suggested in PRECIS to relieve the searcher from the tedium of following the step-by-step references in an index which contains only an entry for the last term in the hierarchy, known as the Bypass routine, has been found disadvantageous for conducting generic searches (Weintraub, 1979, pp. 113-115).

One of the principal functions of a PRECIS-type thesaurus is that, it allows the use of shorter and simpler index strings without loss of access points. Terms that are clearly present in a topic by implication (to the extent that we imply some kind of 'cartographic materials' whenever we speak of 'maps') can usually be excluded from index strings and index entries, and handled instead by once-for-all references as in PRECIS:

Cartographic materials

See also
Maps

But there are certain drawbacks to adding this second dimension (i.e.,
thesaural relations or organizing classification effect) to the more obvious first dimension of syntactic relations (i.e., associative classification effect) in an indexing language. This produces index strings having large number of components in them, as in DSIS. User surveys have found that subject strings with eight or more component terms present difficulties in interpretation (Raghavan and Iyer, 1978). The problem raised by the DSIS's proposal to modulate modifier of kind 1 forming complex term has been pointed out by Biswas and Smith (1988a, p. 7; also see Section 6.1.3, Chap. 6), and can be further substantiated from the findings of the users' survey mentioned above, which concluded that "Introduction of subject field terms and superordinate links in a subject string where they are superfluous leads to confusion among information system users and therefore possible misinterpretation of the subject strings. It is, therefore, suggested that the subject field terms and superordinate links be omitted in structuring a subject heading except when these are necessary for clarity (e.g. as a context indicator)" (p. 12). Thus, we can suggest that indiscriminate use of hierarchical relations in the index string itself is rather detrimental to the searcher's understanding of the entries and pushing the break off point in a search of the irrelevant entry further away. Besides, it increases the index bulk considerably, making the search process tedious and inefficient.

It is entirely possible and, in fact, occurs fairly often that a concept has two or more broader concepts. The relation between the concept and its two or more broader concepts is said to be polyhierarchical (ISO 2788). This may apply to both the generic and hierarchical whole-part relationships. For example,
But the standards suggest no way of defining the different hierarchies, and no cross-link is made between say 'Audiovisual aids' and 'Instructional materials' in the first example [12] even though these two terms are really 'overlapping' terms (from variant hierarchies) and each could divide the other. The existence of this problem is also confirmed by Willetts (1975), in her study of the relation between terms in thesauri. Wall (1980) has proposed that overlapping terms should be determined by the 'some-some' test as follows:

![Diagram of relation between overlapping hierarchies](image)

Figure 7.3: Relation between Overlapping Hierarchies

and should be designated as 'XTs' in relation to each other, instead of the standard 'RTs'. PRECIS resolves the problem of polyhierarchies successfully by generating multiple 'descending' See also references, such as
Devadason (1985b, p. 16; Devadason and Kothanda Ramanujam, 1982, p. 174) has also agreed to the existence of polyhierarchies in the Classaurus when he says "like a thesaurus any term is permitted to appear in as many hierarchies as may be appropriate". In the context of DSIS entry generation he suggests that "If it is identified that the particular component term being chosen to form the context is polyhierarchic (ascertained from the alphabetical chain index to the concerned Classaurus), then the successive superordinate terms to it that resolve the homonym should also be selected to form the Context" (Devadason, 1986a, p. 17). This certainly will help to solve the situation faced by the indexer in the second example above, but not the one in the first example, where the last concept 'Instructional films' is linked to the broadest concept 'Educational media' through two different intermediate broader concepts, viz., 'Audiovisual aids' and 'Instructional materials'. It is quite possible that some documents dealing with certain types of 'Audiovisual aids' (e.g., Instructional films, Protocol materials) can as well be useful for someone interested in similar types of 'Instructional materials' and vice versa. On several occasions the indexer might be faced with the dilemma of making a choice between more than one equally preferable alternative hierarchies. Similar to the failure of standards in suggesting any means of cross-link between overlapping terms, Classaurus does suffer from the same drawbacks. This problem is to some extent similar to the one created by the system's failure to recognise the fact that a modifyee (focus) can be modified by two or moreModifiers (differences) on the same level and thus give birth to polyhierarchies (see Section 7.3.1.2 above).
7.3.2.2 Associative Relationships

The rules of syntax of DSIS requiring the discipline term to be represented as the first context specifying category eliminates the need for *See also* cross references of the 'associative' type 'Learning *See also* Education' and 'Education *See also* Learning', practised in PRECIS (Austin, 1984, p. 209). Attempts to enumerate non-hierarchically related terms of a particular manifestation, such as above, has only created confusions. No two specialists agree in the choice of so-called 'Related Terms'. But if this part of the activity is left to the care of the subject strings themselves, no two specialists can disagree; for, in this process two terms are said to be related because they have occurred as related in the sources of information (Bhattacharyya, 1982, p. 264). DSIS incorporates all these features. Thus, on the whole, DSIS entries collocates better than the PRECIS entries.

7.3.3 Categorical Structure

The debate surrounding what constitutes an array of totally exhaustive and mutually exclusive categories, is a long and controversial one. It starts usually by citing the categories of Aristotle, who gave them an ontological dimension, and finishes with a discussion of categories (based on literary warrant) prescribed by eminent CRG (Classification Research Group, UK) members such as, Vickery (1975), Foskett (1970) and Austin (1984). In between come the works of Kaiser (1911) and Ranganathan (1967). But none of these solutions could be regarded as an absolute one (Svenonius, 1979), each one of them having faltered in 'category definition' in one way or another. Having said so, little positive change can be expected in the conceptualization of the Elementary Categories in DSIS, which subsequently form the facets in the Systematic Part of the Classaurus. Because, as mentioned earlier, these Categories were drawn on the basis of the Deep
Structure of the Subject Indexing Languages (SILs) forming part of the General Theory of SIL, which itself was developed through logical abstraction of the structures of outstanding SILs -- such as those of Cutter, Dewey, Kaiser, and Ranganathan (Bhattacharyya, 1979a).

Let us consider a practical example to illustrate our point of contention. We will consider the name of subject used as an example to explain certain aspects of DSIS syntax in Section 6.2, Chapter 6: 'In education, curriculum for in-service training of non-teaching personnel in polytechnics'. The entry structure suggested by Devadason thereof was as follows:

```
Education 8 Curriculum 8.5 (for) In-service training 8.5 (of) Non-teaching personnel 8.5 (in) Polytechnics
```

The following online Classaurus entries generated by the above suggested input string will reveal limitations of the categorical structure adopted for DSIS, which will subsequently form the basis of the schedule of facets in Classaurus:

```
Curriculum

(for)

—in-service training

(of)

—non-teaching personnel

(in)

—polytechnics
```

[Note: Not arranged alphabetically; all three are Modifiers of Kind 1 to the Entity term 'Curriculum', hence, printed at the same indention.]

But the second entry, i.e.,
Curriculum
(of)
-Non-teaching personnel

is a total distortion of the meaning it was supposed to convey. Instead, an entry like,

Curriculum
(for)
-Non-teaching personnel

would have been more logical and appropriate in this case. One can easily visualize the would be Classaurus entries and resultant confusion generated by the second input string provided in Section 6.2, Chapter 6 (i.e., one minus the 'Curriculum' concept) very well. For example, it was indexed (following Devadason's procedure) as:

Education 8.1 In-service training 8.1.5 (of) Non-teaching personnel
8.1.5 (in) Polytechnics

and generate Classaurus entries (again, not arranged alphabetically) as:

In-service training
(of)
-Non-teaching personnel
(in)
-Polytechnics

Therefore, it seems to be logical to conclude that the number of ECs postulated for DSIS and Classaurus is not adequate enough to deal with every possible concept in the name of a subject (cf. Section 7.3.1.2 above). However, one might ask "Is the agent of action a category of existence like a part of thing? Might it not be more logical to regard agent as a
grammatical (deep case) category rather than as a paradigmatic one?" (Svenonius, 1979, p. 70). Quite right. The 'Curriculum' can be an Entity as well as the agent (instrument) of an Action such as 'In-service training'. The former's mode of being is essential (semantic), whereas, that of the latter is accidental or contingent (grammatical). In this sense, the action/process term 'In-service training' is associated with a set of entities possessing different roles (or deep cases) such as,

- Polytechnics = whole (containing the patient)
- Non-teaching personnel = patient
- Curriculum = instrument,

and these should not be conjoined in a hierarchical genus-species relationship. Following Austin (1982, pp. 90-91), we can say that, a hierarchy such as the one resulting from Devadason's input string "should be rated as logically unacceptable on the grounds that concepts of such radically different kinds cannot share roles as coordinate members of a class" of entities (as instruments). In the proposed system of categories in the Classaurus, there is bound to be a mix up of semantic and grammatical categories. The result of which is serious, viz., the non-mutual exclusivity of categories. It is hard to believe that Ranganathan himself (had he been alive) would have agreed to such aberration in the treatment of terms which act as the building blocks of the system. In defining facets (categories), Svenonius (1978, p. 141) observes that "if categorization or classification of terminology is introduced for a systematic purpose, such as information retrieval, care must be devoted to definitions. Categories must be well defined in the sense that conditions for membership are explicitly stated". This process can effectively be halted by making provision for additional synthetic devices for input coding in the DSIS, which will take care of representing its grammar alone. Making grammar (syntax) rigidly dependent on semantics invites more problems than it solves. Of course, each can
complement the other, while still remaining separable.

Similarly, a number of authors (Langridge, 1976; Svenonius, 1979) have criticised the way certain categories have been named and defined in PRECIS. One such controversial category deals with concepts describing 'phenomena', such as, 'football', 'disease', 'foreign relations', etc. PRECIS treats such terms as actions and assigns operator 2 to designate them (DSIS as well as Colon classification would treat 'disease' as a Property or Matter-Property category concept). A term is a phenomenon term if (i) "it appears to represent things engaged in action rather than an action per se and (ii) it cannot be reduced to an infinitive (Austin, 1974, p. 135). The mixture of this dual criteria (semantic and morphological) for category definition has been criticised by Svenonius as some sort of compromise by Austin. Langridge (p. 211) questioned the justification of categorizing concepts such as 'emotional development', 'academic achievement', etc., as actions, while 'attitudes' as property.

7.3.4 Need for Proper Structure in the Systematic Display

Like a faceted classification scheme, Classaurus consists of separate schedules for each of the 'facets' or Elementary Categories (Fundamental Categories and NOT characteristics of division like 'by shape', 'by wavelength', etc.), viz., Discipline, Entity, Property, and Action; with their Species/Parts and Special Modifiers for each (Devadason, 1985, p. 16). This is quite understandable. But it should be remembered that "Hierarchy is not a straightjacket in which the universe of knowledge has to fit somehow or other. On the contrary, a properly designed hierarchy is a device to assist in indexing and in performing searches. Whenever a hierarchy sets constraints, it is faulty" (Soergel, 1974, p. 78). The proponents of Classaurus seem to have forgotten this. Even within a single facet or
Elementary Category, concepts are liable to be divided into further sub-facets on the basis of several characteristics each applied at a time. Strictly faceted classification schemes such as Ranaganathan's (1960) Colon Classification contain numerous examples of facets subdivided according to characteristics of division. For example, the Personality facet 'Social group' in the Main Class 'Sociology' is subdivided according to the following characteristics:

1 By age and sex
2 (By) Family
3 By residence
4 By occupation
5 By birth or status
etc.

Similar display of further subgrouping of individual facets preceded by facet indicators in an experimental Classaurus of political science using computer has been reported by Sethi and Shyamala (1981/82, pp. 118-120). "In alphabetical displays it is usual to mix, in one sequence, subordinate terms characterized by different principles of division, whereas in systematic displays it is possible to achieve helpful subgroupings of homogeneous terms at the same hierarchical level. The subgroups are preceded by facet indicators (named 'node labels in ISO 2788')" (Aitchison and Gilchrist, 1987). The PRECIS thesaurus also makes provisions for including such node labels, especially when there are several 'orphan' terms within the same subject area (Austin, 1984, p. 273). But, for unspecified reasons, such facet indicators are excluded from the Systematic Part of the online Classaurus by Devadason (1985, p. 22). May be it is the difficulty of arriving at any other logical sequence other than alphabetical sequence by using computer, which has forced the originator of the system to use the principle of alphabetical sequence in arranging the terms not only in an
array but also the sub-facets within a single Elementary Category. The use of classificatory techniques gives several benefits. Classification shows missing hierarchical levels and may also show gaps in hierarchies. At this stage of our study, it is not clear how much of the benefits of a classed vocabulary such as Classaurus we can barter for the sake of its computerization. However, the benefits of such an arrangement are bound to be minimal in comparison with those of the other classed vocabularies, since it has been already recognized that "thesauri without formal structures (at all levels) are no real help for the indexer" (Dahlberg, 1988, p. 1).

7.3.5 Need for Special Codes in DSIS

The following cases describe some of the pressing situations where we felt that the prescribed coding conventions were inadequate to produce desired results in the Classaurus entries.

7.3.5.1 To Indicate Coordinate Concepts

In Section 6.2.2, Chapter 6, we have the problems of handling coordinate concepts in DSIS and the implications of the solutions suggested by Devadason for the index generated. For example, subjects like 'Installation of microfilm readers and photocopiers in public libraries', DSIS proposes to treat as 'multi-focal or multi-theme' documents, which would require separate names of subject for each theme. The failure of such proposal, especially for what Austin (1984, p. 97) calls the 'Bound coordinate concepts', is pointed out by Biswas and Smith (1988a, pp.5-6). However, here we would like to discuss those 'special situations' where the conjunction 'and' may also be used in a Complex Term to form multifocal component (Devadason, 1986a, p. 15), such as:

Teaching 8.1.5 (using) Interactive videos 8.1.5 (and) Microcomputers.
Unless the second Modifier 'Microcomputers' is preceded by some sort of code other than the EC indicators and the prescribed set of process codes, it would generate the following set of Classaurus entries given on the left-hand side below, instead of the one on the right-hand side:

- Teaching (using)
- Interactive videos
- Microcomputers

(And)

Teaching (using)
- Interactive videos
- Microcomputers

Needless to say, the set of Classaurus entries on the right-hand side above, is akin to the one which Devadason might have favoured.

7.3.5.2 To Identify Same Parts for a Term and Its Species

As pointed out by Devadason (1985, p. 25) at the end of his paper that "Another major problem is the appearance of the same Parts for a term and also for a term denoting its Species", we also felt the same. For example, 'Catalogues', 'Shelves', 'Books', 'Serials', etc. may constitute the Parts of the 'Libraries' as well as many of its Species such as, 'Public libraries', 'School libraries', 'Lending libraries', etc. Devadason mentioned a program called CHECK, being developed to analyse the Alphabetical Part of the Classaurus and to point these out. Alternatively, we suggest that such cases can be dealt with by an additional code which would identify input strings containing such common Parts of a term and also of its Species, and jumps one step to link the Part with the Whole rather than its Species. As a result, the terms representing Parts of a term will be printed after the term itself and will mean that these Parts are equally applicable to all of its Species. Of course, those Parts which only belong
to some specific Species, will be printed below them. However, we are yet in a position to test the implication of this proposal thoroughly.

7.3.6 Potential as a Multi-disciplinary Vocabulary Control Tool

There has been very little use of POPSI (and DSIS) in practice, except the following (Vinayak and Taneja, 1986):

2. *Mohandas Karamchand Gandhi: A Bibliography*. Indian Council of Social Science Research, New Delhi, India, 1974

Consequently, there has been little application of Classaurus as a vocabulary control tool as such. It has been mainly used for the purpose of in vitro laboratory experiments involving student project works at DRTC, Bangalore. Starting from December 1979, attempts are being made to introduce POPSI in a limited way in the subject indexing operations of Jawaharlal Nehru University Library, New Delhi, India (Sethi and Shyamala, 1981/82). Classauri have been designed for some of the social science subjects like economics, political science and sociology, and applied to batches of 100 books and articles in each subject. The results have been claimed to be fairly satisfactory, though much remains to be desired on its computerization. We have already expressed our reservation about DSIS' potential vis-a-vis PRECIS with respect to the generation of a multi-disciplinary index such as *British National Bibliography* subject index (Biswa and Smith, 1988a, p.10; also see Section 6.2.1, Chap. 6). Besides the possibility of enormous increase in the number of disciplines (e.g., the 7th edition of *Colon Classification* has an incredible list of 776 Basic Subjects (Dhyani, 1988), most of which can be treated as Disciplines in
terms of DSIS), the necessity for provision by multi-disciplinary approaches to the same or similar topics in a general index happens to be the least cost effective, since, it consumes more input time and output space. Accordingly there have to be as many Classauri as there are disciplines, which again are ever changing. Classaurus might prove to be a useful and economic tool in a system involving a single discipline or two, but it will certainly be uneconomic for the production of a general vocabulary involving the universe of knowledge.

Notes

1 Compound Subject -- "A subject with a basic subject and one or more isolate ideas as components".

   Compound Class -- "A Compound Subject taken along with its rank, as fixed in the course of the successive assortments of the Universe of Compound Subjects" (Ranganathan, 1967, p. 84, 86).

2 Complex Subject -- "Subject formed by coupling two or more subjects expounding, or on the basis of, some relation between them".

   Complex Class -- "A Complex Subject taken along with its rank, as fixed in the course of the successive assortments of the Universe of Complex Subjects" (Ranganathan, 1967, p. 85, 86).

3 Writing in 1970 Ranganathan (Ranganathn, 1970, pp. 17-18) preferred to call it a 'Compound Subject Term' rather than the old 'Compound Class Term', which "recognises the fact that a subject qua subject is not a class. It becomes a class if and only if it is 'ranked' and given a definite place in the sequence of all the subjects".

4 "Any idea or idea-complex fit to form a component of a subject, but not by itself fit to be deemed to be a subject" is an Isolate or Isolate Idea (Ranganathan, 1969, p. 200). "The term Isolate is applicable equally in the Plane of 1) Idea, 2) Language, and 3) Notation... In the Plane of
Language (Verbal Plane), 'Isolate Term' is the equivalent of 'Isolate'" (Ranganathan, 1957, p. 1.58).

5 "Two or more isolates or an isolate and a special component can combine to form a compound isolate", where "A special component is an idea, which is not by itself a subject or an isolate, but which can be used as a component to be attached to a host isolate as well as to its subdivisions, in order to form a compound isolate" (Gopinath, 1976, p. 56). For example, 'Steel pipes' is a compound isolate, the 'Steel' and 'Pipes' being two independent isolate ideas. Again, 'Old English' is a compound isolate idea, having the isolate 'English' as the principal component; the idea 'Old' is a special component formed on the basis of the characteristics 'by stage'. "The generic term 'Compound Isolate' is used to denote Compound Isolate Idea, Compound Isolate Term, and Compound Isolate Number" (italics mine) (Ranganathan, 1970, p. 19).

6 "It is possible to have books (documents) expounding the relation between two isolates in one and the same facet of a class. We shall call it 'Intra-Facet Relation'. An isolate formed by thus bringing into relation two isolates in the same facet shall be called a 'Complex Isolate" (Ranganathan, 1960, p. 1.57). In the Plane of Language 'Complex Isolate Term' is the equivalent of 'Complex Isolate'.

7 It is being reported that the 7th edition of Colon Classification -- basic and depth version, volume I, has finally been published (Dhyani, 1988). Instead of DE, chapter DG now enumerates language isolates (Ranganathan, 1987).

8 "Concepts that cannot be decomposed (split) further in the [a] given system... are called 'elemental concepts'" (Soergel, 1974, p. 74).

9 More simply "A single or multi-worded term and representing a compound concept" (italics mine) is a Composite Term (Seetharama, 1976, p. 67).

10 Classaurus is necessarily the indexer's tool. Since, all the hierarchical and associative relationships among terms are present in the DSIS index.
entries, there is no need for the user to consult the Classaurus.

11 A 'Quasi Isolate Idea' represents the various 'characteristics of
division' or 'facet indicators' or 'node labels', and a 'Superimposed
Isolate Idea' is formed by connecting together two isolate ideas
belonging to one and the same facet (Ranganathan, 1967, p. 425).

12 This example has been drawn from the Thesaurus of ERIC Descriptors
(1987), which, of course, provides cross-link (through 'RT' relation)
between 'Audiovisual aids' and 'Instructional materials'.
References


Reserve University.


Devadason, F. J. (1986b). Personal communication.


London, G. (1965). A classed thesaurus as an intermediary between textual,


As mentioned in Section 5.3.1, Chapter 5, one of the objectives of the present study was to find out the respective searching performances of PRECIS and DSIS indexes, the present chapter will try to throw some light on this aspect of these two string indexes. It must be remembered that, the very idea of string indexes has originated from the need to supply the users with a tool, which can provide access to the same subject from whatever point the user may decide to approach the index. Thus, it might be an useful venture to have an assessment of how effectively and with what efficiency these indexes operate. Search effectiveness is usually measured by what is popularly known as the 'recall' and 'precision' ratios. Three distinct characteristics which distinguish a string index from other types of index are the presence of: (i) multiple index entries, (ii) overlap among entries, and (iii) explicit syntactic rules. All these characteristics are aimed at improving the efficiency and effectiveness of searches.

The aim of multiplication of index entries is to increase efficiency and the recall side of the effectiveness by giving searchers increased direct access to useful information. Overlap is calculated to improve both recall and precision through increased detail in index entries. Increased detail improves precision by helping the searcher to decide the item's usefulness. Recall too is improved, because the searcher may incorrectly assume that an inadequately described item is not worth retrieving. The explicit syntactic rules, on the other hand, are designed to increase recall and efficiency and, to some extent, precision. The various ways that these rules accomplish these purposes, have been identified as the five major characteristics of index strings from a searcher's point of view (Craven, 1986, pp. 6-11) [1]:
(1) **Predictability** refers to the extent to which the searcher can predict the forms that relevant entries will take and, more especially, where they are likely to be found in the index display.

(2) **Col location** means the placing of similar index entries together and the separation of dissimilar index entries.

(3) **Clarity** refers to how likely it is that the entry will not be misinterpreted and to how readily searchers can correctly grasp the meaning of it.

(4) **Succinctness** represents the ratio of detail or specificity to the length of the entry.

(5) **Eliminability** means how quickly a searcher can break off examining an index entry if it is irrelevant.

Some of the above characteristics may counteract with each other. For example, excessive succinctness may detract from clarity. Apart from these, Craven mentioned some other factors like redundant information, irrelevant information and excessive index bulk, which waste searcher effort and so decrease efficiency. "**Redundant** information is information which the searcher already has; **irrelevant** information is information which the searcher does not particularly need" (p. 9). For the purpose of this study, the aim is to assess and compare the performances of PRECIS and DSIS in terms of the above-mentioned characteristics and on a non-empirical basis considering it from the viewpoint of potential searchers of the indexes. This tentative comparison of searching is basically confined to the respective printed formats produced by both the indexing systems.

The operation of searching may be performed by a person from different angles or 'modes'. Keen (1977b, p. 269) has identified these as:

1) **Seeking mode** -- that of finding a sequence of entries which looks likely to contain useful information.
2) Scanning mode -- that of looking at the more prominent features of such entries, such as their access points, in order to select those entries which still seem likely to be useful.

3) Screening mode -- that of examining selected entries in detail to make final decisions on which indexed items to retrieve.

Different indexing systems try to match these three different modes in different ways and apply different weighting to respective parts of index entries they produce. The three different functions, viz., 'locating', 'comprehending' and 'relating', which subject index entries are supposed to perform (also identified by Keen, 1977a, p. 19), try to match the above-mentioned modes of search in an amalgamated manner. The locating and relating functions try to match mostly the seeking mode and to a lesser extent the scanning mode. The comprehending function tries to satisfy the screening mode, and to a lesser extent the scanning mode. Similarly, different qualities of index strings such as, predictability, collocation, eliminability, etc., tend to aid to different extents in different mode of searching. For example, predictability is especially important in the first mode; collocation and eliminability are important in the second mode; and clarity is important in the third mode.

Among the earlier comparative studies carried out between PRECIS and POPS (reviewed in Section 5.2, Chap. 5), none has attempted so far to compare the performance of the respective systems from the users' point of view. So far as PRECIS is concerned, Keen's EPSILON project (Keen and Wheatley, 1978; Keen, 1978), the Wollongong University Subject Catalogue Study (WUSCS) (Hunt et al, 1976/77), and the Liverpool Polytechnique study (Peters and Bakewell, 1981) assessed the users' preferences and/or reactions to PRECIS indexes. EPSILON and WUSCS looked mainly at search efficiency and effectiveness and searches' preferences for indexes to
small databases. The findings of both these experiments were based on actual searches carried out in laboratory and operational environments, respectively. The Liverpool Polytechnic investigation tried to assess the impact of PRECIS on the users of published PRECIS indexes and those operational PRECIS indexes in individual libraries. The findings were mainly in the form of subjective opinions collected through questionnaires and interviews. PRECIS also figured in another small searching test conducted at Loughborough University (Scardellato, 1979). On the other hand, POPSI has attracted only one users' survey till this date (Raghavan and Iyer, 1978). This conducted detailed surveys of users' ability to interpret POPSI subject strings in the field of social sciences.

8.1 Terminology

8.1.1 Form of Terms

In ordinary language a single term may be used to represent various meanings (homographs) or different terms may be used to represent same meaning (synonyms). Also, the orthography (e.g., 'online' or 'on-line') and form (i.e., singular vs. plural) of a term may vary according to the culture and the context, respectively, in which it is being used. But in a controlled language such as an indexing language, it is of utmost importance that terms and their use should be standardized. Some of the criteria on which terms are chosen include unambiguosness, familiarity to users, brevity, and relevance in distinguishing and grouping indexed items. For index entry displays factors such as, collocation and predictability of filing must also be considered. In string indexing choice of terms is also partly determined by the layout of the index strings.

Both PRECIS and DSIS use the context principle, i.e., the presence of other terms in the string, to resolve the searchers uncertainty about which
meaning is intended. For example, a subject such as 'Periodicals in libraries' would be coded in PRECIS as:

*(1) libraries
(p) stock
*(q) periodicals

rather than as:

*(1) libraries
*(p) periodicals

The extra term ('stock') in the first string helps to avoid the possible confusion on the part of the user which might have arisen from the second entry generated by the second string:

- Periodicals. Libraries

To many people this might appear to be meaning 'Periodicals about libraries'. Similar treatment is recommended in DSIS which uses successive superordinate terms as contexts to resolve the homonyms. But, as we have seen in Section 7.3.2.1, Chapter 7, this technique fails to provide a hundred percent satisfactory output in the case of polyhierarchical terms having a common superordinate top link.

In the case of PRECIS, the class to which a concept belongs generally determine the form of its term in the index: in particular, its expression as a plural or singular noun. These are set down in the form of a table which follow the recommendations of BS 5723. To show the fact that such rules are applied by human beings almost intuitively, Austin (1984, p. 106) quoted Lewis Carroll, who for example talked "...of many things --"

"...of shoes -- and ships -- and sealing-wax --

Of cabbages -- and kings --

And why the sea is boiling hot --

And whether pigs have wings".

-- where all count nouns are expressed as plurals, and the non-count nouns
as singulars. This is necessary for the sake of consistency in index displays. For DSIS no such rules were provided. The only mention of consistent use of terminology is in relation to auxiliary/function words forming complex terms. All the examples of noun words in the index strings given in the FID/CR Report (Devadason, 1986a), in fact, are in their singular forms. In his correspondence with this author, Devadason (1986b) rather satirically quoted the following lines from a poem published in Readers' Digest:

"Box, boxes; ox, oxen?
Brother, bretheren; sister, sistren?"

However, his final recommendation was to use the respective national or international standards (if any) as guidelines for constructing indexes in the particular language concerned. Following his suggestion this study used the recommendations of BS 5723 as guidelines, as it is being used for PRECIS.

8.1.2 Access Terms

In Section 6.3.1, Chapter 6, we have seen the nature and implications of the constraint imposed upon the indexer by the rules of DSIS in choosing lead terms, especially when some insignificant terms form the first constituent of a Complex Term. The practical implications of such rendering for the searcher is two-fold. First, he has to tread across a number of unsought headings which he thinks useless and irrelevant. Secondly, when the second and subsequent constituents of a complex term form the lead, the entries become indirect, i.e., (permuted) cross reference entries involving two-step searches. The first factor is detrimental to the efficiency of the index whereas the second factor may lead to irritation on the part of the searcher. Our study has shown that approximately 1/10th of the entries (310 out of a total of 3074 entries generated from 600 strings) in the DSIS index
were permuted cross reference entries requiring leads under subsequent terms in a Complex Term. In comparative terms, these entries have the same effect as those See also entries in PRECIS generated to link the prepositional phrase with other entries having leads under the modifying noun in such a phrase. For example,

PRECIS: Children
   See also
   Libraries for children

DSIS: Children / Libraries (for)

Compared to this, the PRECIS indexer can generate or suppress access under any word in the string by inserting the appropriate codes. As a result, it can be said that, the searcher using such an index is about to experience less noise in searching (Biswas and Smith, 1988).

8.1.3 Index Term Context

The provision of index term context is an integral part of the index entry. This becomes more essential if the index is a pre-coordinated one. The amount of context to be included in index entries is a debatable issue. It varies from one system to the other. For example, at one extreme, systems such as PRECIS provides full context under each lead term, while on the other extreme, the index to Current Index to Journals in Education uses no context at all. In between come a large number of indexing systems using variable contexts, e.g., chain index. A question often asked is 'how much context' should be called optimal (Svenonius, 1982). In the case of PRECIS index, the context is determined mainly by the total number of components in the name of a subject which are related on the syntactical axis, with occasional exceptions of extra words inserted to resolve any ambiguity that may arise or to provide better collocation in the index displays. The use of
operator q can be named specifically as performing the latter type of function. The responsibility for showing the dimension of semantic relatedness between terms is delegated to the thesaurus. On the other hand, in Devadason's (1986a, p. 17) opinion, apart from their primary role of helping the user in comprehending the subject index entry and to permit relevance prediction, "the Context Heading (in DSIS) also performs the 'relating function' by providing an organizing sequence among the Context Headings to a particular Lead Heading when just sorted alphanumerically. In order to get the maximum possible organizing effect, all the component terms in the name of subject including those interpolated/extrapolated at the 'modulation' step, formalised and arranged according to the rules of syntax, along with the indicators ... should be kept in the Context Heading". In simple terms, the DSIS index entries should incorporate within the frame of a single index entry both the dimensions, viz., syntactic as well as semantic. The practical benefit for the searcher of a PRECIS index is that he is faced with relatively shorter index strings, and if his need is specific and can be satisfied with fewer relevant documents, then PRECIS index is an advantage. On the contrary, the DSIS index (Lead heading with upper link specifiers and full context) might help a searcher needing an elaborate search, for example, a researcher looking for retrospective materials on his topic of research. In the case of a PRECIS index this might prove a very frustrating experience, as he has to run through a maze of 'descending'. See also references. But at the same time, as we have seen in Sections 6.1.3 and 6.3.2, Chap. 6, the searcher may fail to comprehend the meaning of the name of subject from a DSIS string which is both long and overly time-consuming to read. In Craven's (1986, p. 130) opinion, this might prove to be a hindrance to searchers.

In addition to the superordinate terms forming part of the DSIS context headings, the indicators of deep structure (same as the role operators in
PRECIS) are also included in it to increase the organizing capability of the context headings. But no conclusive proof has been provided so far that such an alphanumeric arrangement will be liked by the users, not to mention of their ability to understand and benefit from such an arrangement. From an aesthetic point of view, the DSIS index will certainly be ranked lower than the PRECIS index which uses basically a fragmented term order with occasional use of function words.

8.1.4 Compound Terms as Lead and Context

In Section 6.1, Chap. 6 and Section 7.3.1, Chap. 7, respectively, we have observed the type of leads in index entries and representation of terms in vocabulary control tools, that will result from PRECIS' and DSIS' treatment of compound terms. However, the DSIS index user is not concerned with the Classaurus at all, since Classaurus is solely the indexer's tool. We have also seen that a DSIS index fails to provide hundred percent recall in those cases where a compound term is modified with more than one modifier semantically related to the focus on the same level. PRECIS prefers to repeat the compound term in the display until the whole term appears in the lead, which gives rise to certain amount of redundancy. For example, a compound term such as 'Adult vocational education' could be input as:

*(2) education $21 vocational $21 adult

and would generate the following entries:

Education

Adult vocational education

Adult education

Adult vocational education

Vocational education

Adult vocational education
Such repetition might be considered as redundant. PRECIS, in comparison with some other indexing systems (e.g., Relational Indexing), weighs the disadvantages of redundancy relatively lightly against the advantage of clarity and prefers to repeat the modified noun, as shown above. These entries give better precision by revealing to the searcher that the document indexed is specifically on 'Adult vocational education', which also makes it easier for him to break off reading at an early stage if he is not interested in it. DSIS repeats the whole hierarchy of superordinate links in the context, but fails to collocate it under 'Adult education' as well, as shown below:

 STRING:  education $3 vocational $3 adult

 ENTRIES:  Education

 Education. Vocational Education. Adult vocational education
 Vocational Education
 Education. Vocational education. Adult vocational education
 Adult vocational education
 Education. Vocational education. Adult vocational education

[Note: Entries are rendered without the indicators.]

Such entries might provide overall collocation better, e.g., under the top superordinate term 'Education', but not under all of its species. Also, apart from difficulties of comprehension, the searcher has to wait till the last term in certain entries (e.g., the entry under 'Vocational education') to decide finally that he is not interested in 'Adult vocational education'. Thus, eliminability is worse.

The other category of compound term (complex term in DSIS), namely, prepositional phrases (plus terms with auxiliary/function words in between in DSIS), appeared more frequently as leads in DSIS than in PRECIS. There were 383 such lead headings in a total of 3074 entries in DSIS index.
(approx. 12.46%) against only 22 such leads in a total of 2583 entries in PRECIS index (approx. 0.85%), which suggests that lead terms in PRECIS are more succinct than in DSIS. Among those 383 lead headings 180 had prepositions in them (approx. 47.0%). The remaining 203 had other auxiliary/function words in them (approx. 53.0%), most of which belong to the category of phase relation indicating phrases.

8.2 Single-entry vs. Multiple-entry Citation Order

The principles of citation order were initially developed to fulfil the need of systems which require only one order of components to retrieve a single item. The library classification systems belong to this genre. The need for citation order principles becomes more acute in string indexing systems, where most of the significant components in the input string act as lead term of one or more index entries. Still most of the string indexing systems are based on one citation order in the input string. Even PRECIS was initially developed on the basis of a single-citation order and still now majority of index strings are generated using the 'standard format'. The other two formats were added later to clarify the roles of certain terms in the index string in relation to other adjacent terms and to achieve helpful collocation in entries with certain types of term as leads. Principles of citation order actually refer to general rules which are supposed to contribute to good qualities in index strings such as those mentioned at the beginning of this chapter, which in turn enhances the efficiency and effectiveness of searching. Ranganathan could be named as the founder of the one of the most well-known theories of single-entry citation order, viz., that of PMEST, which he considered might represent an 'absolute syntax' of subject propositions. He also developed a number of supporting principles, such as Wall-Picture principle, Cow-Calf principle, etc., for deriving a syntax of facets parallel to the syntax of ideas in human mind. The order of
The principle of context determines that, a term should be followed by those other terms which help most to narrow its scope in such a way that the searcher is able to comprehend its meaning. Indexing systems using context principle mostly do so, because it helps to increase the overall clarity of the index entries. Both PRECIS and DSIS prescribe that the full subject analysis should be represented in each entry to render an index entry coextensive with the subject of the document being indexed. Its other role is to increase eliminability, by helping searchers to decide the break off point at an early stage of an entry which he thinks irrelevant. But, as we have seen in the earlier chapters that, DSIS' proposal to modulate the manifestations of each EC by interpolating/extrapolating with superordinate links, may lead rather to loss of comprehension and decrease of clarity. For example, the following two sets of index strings deal with the same type of facets in the PMEST formula were derived on the basis of what Ranganathan calls the principle of 'decreasing concreteness', which states that less concrete, more abstract, terms should be cited after more concrete terms. The 'deep structure' of POPSI (and DSIS) is derived from this facet syntax, which is artificially postulated. On the other hand, the citation order principle used by PRECIS, viz., the 'context-dependency' principle is regarded as probably the most well-known principle for arranging concepts in string indexes. Even Devadason (1986a, p. 6) approves of it and hypothesizes by saying that "the rules of syntax (of DSIS) give rise to a context-dependent sequence of the components in the name of subject in conformity with Ranganathan's Principles for Facet Sequence" (italics mine). The context-dependency principle has been seen by Craven (1986, pp. 96-98) as a combined manifestation of two separate principles of 'context' and 'dependency'. The first of these principles determines what terms are placed together in a citation order and the second establishes which of two adjoining terms should be cited first.
subject statements, but clearly vary on the level of clarity:

Set 1:
Synopsis journals. Online acquisition
Libraries. Stock: Synopsis journals. Online cataloguing
Applications of microcomputers.

Set 2:
Technical processing. Cataloguing. Online cataloguing

In the case of the first set of entries the searcher can easily find out which one of the entries deal with the subject 'Online cataloguing of synopsis journal' and which ones are not. The PRECIS index favours this type of context. But in the case of the second set of entries, it takes longer for the searcher to decide the break off point in an entry if he thinks that it is irrelevant. DSIS principally conforms to such context designations.

The dependency principle, on the other hand, says that the less 'dependent' of two related terms should normally be cited before the more dependent. The basic citation order of both PRECIS and DSIS is geared to generate this order of dependency. But as we have shown on at least one occasion (for the sake of brevity we will mention it as the 'curriculum' example), the rules of syntax in DSIS fail to generate a logical order among concepts in the entry. This has resulted due to the fact that a verb can relate several entities each one of which have a different role to perform.
However, it is not an easy matter to decide whether 'curriculum' should come first followed by 'in-service training' or vice versa. To take a decision in such a case other theories of citation order could be invoked, for example, Coates' 'first coming to mind' principle, Sørensen and Austin's (1976) 'time of conceptualization' principle, etc. According to the later principle, terms in a string should be organized as a sequence according to their relative time of conceptualization as determined by their roles. That is to say, a concept cannot be designated as an agent or instrument (e.g., 'curriculum') unless the action (e.g., 'in-service training') in which it is engaged, whether stated or implied, has already been conceded (to use Ranganathan's terminology). So the order should be 'In-service training, Curriculum'. The same result could be achieved by applying Coates' (1960) link type 15. The main contribution of the dependency principle in index strings is the enhancement of eliminability. The more dependency principle is applied, the less any term in a string has to depend for its meaning on the terms that follow it. But in the case of Devadason's proposed string for the document in question, e.g.,

Curriculum (for) In-service training (of) Non-teaching personnel... etc.

the term 'Curriculum' means very little to the searcher unless the other terms following it were considered. Besides, DSIS' failure to conform to the principle of dependency, the above string also violates the principle of 'Reverse rendering', i.e., the reverse of the sequence of component terms arranged according to the rules of syntax, in the rendering of the Lead Heading. The implication for the searcher is that he has to face another redundant entry. Besides, such practice may prove detrimental to predictability. For example, if another document on say 'In-service training of non-teaching personnel' is rendered as:

In-service training (of) Non-teaching personnel
then it becomes difficult for the searcher to predict what forms the other index entries will take and hence to decide what to look for. It is almost essential for faceted indexing languages such as DSIS, to follow a single citation order within a particular discipline so that predictability of the representation of concepts and concept relations is assured. According to Fugmann (1982), it is inherent in any controlled indexing language that it establishes 'representational predictability' and, hence, prevents serious loss of relevant information, which would otherwise occur in retrieval.

In the case of PRECIS, as we have seen in chapters 2 and 3, respectively, attempts have been made by Austin to establish a correspondence between certain natural language principles and the term order in PRECIS. One such principle suggests that citation order should correspond to the normal order of English passive sentences, that is, 'object-action-agent' (Austin, 1976), although the evidence provided of linguists' support of such word order was rather scanty. In Coates' (1976, p. 94) opinion, the output order of several PRECIS strings follows a bidirectional setting -- the heading following the active order and the display following the passive input order -- which an untrained user may find difficult to comprehend. Keen (1977b, p. 40) has labelled this procedure as 'obscure', which the user could hardly be persuaded to follow in practice.

In certain cases, PRECIS despite its adherence to the principle of citation order mentioned above, uses special coding to reduce the undesirable effects of simple shunting. These are the cases of 'predicate transformation' and 'extra-core operators'. For example, to enhance collocation under terms prefixed with operators 4 to 6 or any of their dependent elements, PRECIS changes to inverted format which ensures that more significant terms are put immediately after the lead terms.
String indexing systems apply control on citation order through various measures. PRECIS is basically rule-based, although many of its features are incorporated in the codes used and in the special routines incorporated in its index string generator. These features bring regularity in the citation order, which provide many advantages to the regular user of a PRECIS index. For example, the searchers may get used to the general pattern of collocation, term order, etc., which they can follow in a rather routine-like manner. But regularity breeds its own brand of difficulties. The resulting rigidity might be a barrier for the user who wishes to enjoy decision-making. DSIS (and POPS1) also has its rules of syntax controlling the overall citation order of terms, but this fixed citation order can be adapted to different types of indexed items and to the needs of different searchers simply by varying the Discipline. However, on theoretical and experimental grounds this might be quite promising, but in practical terms this might prove to be simply a wild goose chase. The amount of effort and time required to establish a list of possible disciplines and the various facets within these schedules of such disciplines could be guessed from an estimate given by Ranganathan's group some years ago and mentioned by Keen (1976, p. 156) was that, to establish depth schedules for subjects in the applied sciences only will take some 5,000 man years.

DSIS' lack of provision for designating certain categories of concepts could be seen by many as a failure to fulfil the need of every type of user. One such category of term indicates the 'viewpoint' from which the subject is examined, for the author's viewpoint can significantly affect the user's relevance prediction. This factor becomes more important for a specialized user who is interested in the minute aspects of his field of interest. The paramountcy of this need (especially in indexing of micro documents) could be exemplified by such contentions as, "indexing which is limited to the
presentation of aboutness serves the novice in a discipline adequately, but
does not serve the scholar or researcher, who is concerned with highly
specific aspects of or points-of-view on a subject" (Weinberg, 1987, p.
241).

8.3 Some other Aspects of Index Strings Contributing to Efficiency

Opposite to the provision of maximum details in index strings, a number of
string indexing systems include some other features which help to shorten
the expressions in the entries without inhibiting the searcher's
comprehension of the meaning. The reason behind their inclusion is thought
to be the enhancement of efficiency in searching or index production. The
PRECIS index string generator includes provisions for omission of certain
terms from the output index which the indexer thinks as unessential in terms
of a meaningful search. For example, a document on the subject 'Video
PATSEARCH' would be input as:

*(1) patents
*(3) online information retrieval systems $w$ for
   (s) use $v$ of $w$ in
*(3) videodiscs
(sub 4 ↑)(1)
*(q) Video PATSEARCH

and would generate the following entries:

Patents

Online information retrieval systems. Use of videodiscs: Video
PATSEARCH

Online information retrieval systems. Patents

Use of videodiscs: Video PATSEARCH
Consider the last entry, i.e., one under 'Video PATSEARCH', where all other terms above it in the string have been omitted on the ground that these terms are thought to be redundant. A person looking under the term 'Video PATSEARCH' is certainly aware of the fact that it is an 'online information retrieval system for patents using videodiscs'. In the case of the DSIS index such omission is impossible unless one changes the whole format of the index entries to be produced, for example, by changing the 'Lead Heading with upper link specifiers and full context heading' into the 'Lead heading with upper link specifiers and short context heading', which obviously might hinder comprehension in certain other entries. This can be seen as an element of rigidity in the system.

The provision of other similar coding conventions (e.g., Not Up (NU), Not Down (ND) and Lead Only (LO), etc.) in PRECIS have identical effects on the index output, i.e., shorter index entries, which are always better from the viewpoint of index generation, and processing the entries during searching. For example, PRECIS occasionally uses an extra term to provide a useful access point in certain subjects, as follows:

*(1) man
*(p) lung (LO)
*(2) pneumonia

This string would generate an additional entry under the term 'lung', but restricts its occurrence in all other entries, where it would be seen as clearly redundant. Usually cross references are used as an alternative to such 'lead only' terms. On the other hand, DSIS employs all lead only terms
to form the context, resulting into longer displays, but a reduction in the number of times a searcher has to look up a new expression to continue a search.

The other major component of string indexes is the cross reference entries. Apart from the simple cross references (those based on single terms) both DSIS and PRECIS (comparatively more rarely) use multiterm expressions as cross references. But there are certain differences in their rendering. For example, a compound concept like 'Concept formation' is factored in PRECIS and represented through a See reference entry, as follows:

```
Concept formation

See

Concepts

Formation
```

The same subject would be represented in DSIS also in a factored form, but rendered rather differently as:

```
Concept formation = Psychology (D) + Concept (E) + Formation (P).
```

In the case of PRECIS index such multiterm expressions as cross references are favoured on the ground of eliminability and collocation, especially when many cross references start with the same term. These also contribute to clarity when the meaning of the 'referred to' concept(s) is unfamiliar to the searcher (e.g., pH See Concentration. Hydrogen ions). There is a basic difference between the rendering of such terms in the two indexes. In an actual PRECIS index entry, the multiterm expressions appear in the same order as that given in the referred to section of the cross reference entry. Therefore, it becomes relatively easier for the searcher to locate such an entry. But, as we have discussed in Section 6.3.4, Chap. 6, in a DSIS index
these referred to multiterm expressions may not be represented in the same contiguous order in the actual index entries. This can make the search process really time-consuming and cumbersome to follow. The other category of cross reference entries prescribed in DSIS, namely, the permuted cross reference entries, have a two-fold effect on the search process. First, these introduce more two-step, indirect searches for the searcher. Second, on certain occasions they hinder clarity, as could be seen from the following heading:

Soviet Union / Foreign relations (with)

The other important aspect of string indexes which affects searching is the formatting of the index display. As index strings consist mostly of text, a well formatted index display would certainly enhance the readability and comprehensibility of the index to the searcher. Craven has classed formatting broadly into 'layout' and 'typography' of the text on the output page or screen. As far as layout is concerned both PRECIS and DSIS use the same basic layout, i.e., heading followed by display or context. Heading is also divided into lead and qualifier (specifier in DSIS). However, PRECIS uses certain typographic variations to distinguish (rather deemphasize) certain segments of the index entry from the other more substantive components in the string. For example, italics is used for printing components prefixed by extra-core operators (4 to 6) and their dependent elements (such as, p, q, etc.) when they appear in the display. DSIS uses parentheses to separate auxiliary/function words from the words they bring together.

Notes

1 The discussion in this chapter is mainly based on the approach followed by Craven (1986) in his recent treatise on subject indexing.
References


Devadason, F. J. (1986b). Personal communication.


and KWOC: Report of a research project designed to examine the
applicability of PRECIS to the subject catalogue of an academic
library. 3 parts. Wollongong, NSW, Australia: University of Wollongong.

(pp. 142-157). London: Clive Bingley.

Keen, E. M. (1977a). On the generation and searching of entries in printed

Keen, E. M. (1977b). On the processing of printed subject index entries

Keen, E. M. (1978). On the performance of nine printed subject index entry
types: A selective report of EPSILON. Aberystwyth, UK: College of
Librarianship Wales.

by laboratory investigation. Aberystwyth, UK: College of Librarianship
Wales. (BLRD Report No. 5454).

indexes. Liverpool, UK: Liverpool Polytechnic, School of Librarianship
and Information Studies. (BLRD Report No. 5659).

Scardellato, K. A. D. (1979). Articulated subject indexing and current
awareness services: A comparison with KWOC and PRECIS indexing in
Loughborough University of Technology.

subjects in social sciences: A users' survey. International

Sørensen, J. and Austin, D. (1976). PRECIS in a multilingual context. Part


CHAPTER 9

SUMMARY AND CONCLUSIONS

Creation of an endless number and variety of patterns in all spheres of life is an inherent characteristic embodied in the very nature of the human being itself. But existence of an opposite and parallel trend could also be evidenced at the same time. Man from the very beginning of its existence on this planet was always busy in finding a uniformity among those enormous number of patterns created by the nature and man himself that surround him. The outcome of this quest are those various man-made systems (of 'artefacts' and 'mentefacts') that try to explain the behaviour of phenomena and events with which he is both familiar and unfamiliar (in the sense of not yet discovering the truth). As a first step to this, man established all those sign systems called 'languages', through which he can express his feelings to fellow men. But at some stages of this transfer process it was felt that these systems were not adequate enough to have a global communication. In came artificial languages such as those proposed by Wilkins and Leibniz, and international auxiliary languages such as, Esperanto, Interlingua, Ido, etc., although with very little practical implementation and success. Similarly, in the sphere of organization and dissemination of recorded knowledge efforts were limited at first within the boundaries of a single institution, slowly extending to that of a society or nation. Advancement of technology brought the scope of dissemination beyond the national boundaries. This resulted in the search for the effective system of communication with universal acceptability. Various solutions were proposed -- from Melvil Dewey's notational system to its offspring Universal Decimal Classification, from faceted systems (such as Colon classification) to coordinate indexing (e.g., Uniterm indexing), from the CRG's proposal to the development of Gardin's SYNTOL, etc. But again, the idea of truly universal
information language appears to be a mirage, although the quest seems to be never-ending. Austin's PRECIS and Devadason's DSIS are two more additions in the same continuum, both making at least some contributions to our understanding of the process of information communication. Following Hutchins (1975, p. 136), we can say that these "may claim closer approximation to universality in DL (Documentary Language or information language) structure than any earlier indexing or classification system".

As far as the present study is concerned, it cannot be considered to be a complete experimental study, to the extent that it was not possible to implement some of the proposed practical testing defined in information retrieval experiment literature as necessary for such experiments.

From the point of view of the respective syntaxes, both PRECIS and DSIS have shown weaknesses in category definition resulting in overlap between the syntactic and semantic structures. PRECIS has tried to shed its classificatory backgroungd with linguistic explanations, but with a limited success. On the other hand, the major problem faced by DSIS was its attempt to incorporate linguistic features (such as the notion of complex term) into a basically classificatory syntax. From the indexer's point of view, DSIS seems to be limited in a number of ways such as, the choice of lead terms, specification of certain classes of concept (e.g., author's viewpoint, coordinate concepts, etc.), limited repertory of choices for controlling the format of individual index strings within a specific index entry type, etc. Its inclusion of semantic context within an entry composed of syntactically related terms has both advantages and disadvantages. If the topic indexed belongs to a field where taxonomic structuring is logical (e.g., medical and biological sciences) then it can be exploited to broaden searches (generalize meaning) and to explode searches (enumerate meaning), especially
in online retrieval. But in a printed format and also in the field of soft sciences (such as the document collection used for this study), its usefulness is debatable.

On the semantic side of the two systems, the effect of decisions taken in the index string generation become obvious. The way DSIS handles the compound terms (especially its adherence to natural language order of the author or document) makes the resultant hierarchies in the Classaurus disjunctive. Coupled with this, its failure to relate terms belonging to overlapping hierarchies makes it prone to loss of information during retrieval. DSIS' division of terms into three classes, viz., compound terms, complex terms and composite terms, also seems to be doubtful, even in the light of Ranganathan's classificatory principles. Classaurus' viability as a multi-disciplinary vocabulary control tool may also prove to be a major obstacle in the way to its implementation as a general retrieval tool such as national bibliographies and large catalogues.

From the searcher's point of view, the format of a PRECIS index may look more aesthetic than a DSIS index. A visual scan of an index composed of alphabets interspersed with numerals may prove tiresome to human eyes. However, in an online retrieval the alphanumeric arrangement of DSIS entries may help to improve collocation considerably. Once the user gets accustomed to using such an index, he might be able to reap the benefits of the inclusion of both the syntagmatic and paradigmatic relationships between terms in the structure of a single index entry. But a full context heading DSIS index would be worse in terms of eliminability. From the point of view of general representational predictability the single-entry citation order of the DSIS index might be quite beneficial. But in those cases where the user wants to make his own decisions regarding the ordering of concepts, a PRECIS index holds certain advantages over the DSIS index. The limited use
of See cross reference entries referring to factored constituents of a compound term (Composite Term in DSIS) might prove quite cumbersome for the user to follow in practice. In terms of computer storage the following two figures can be taken as the measure of the relative bulk of index generated by the two systems (both without cross references): PRECIS -- 379K and DSIS (without upperlink specifiers, but full context heading) -- 764K.

In Section 5.2, Chap. 5, we have mentioned the technique of weighing choices among competing indexing systems proposed by Craven (1986, p. 173-175) in his recent book on string indexing. Based on this technique Craven presented a tentative comparison among seven string indexing systems considering a total of 8 different features for each system. The various stages of the weighing process along with their respective ratings for PRECIS and POPS were presented as follows:

<table>
<thead>
<tr>
<th>Features</th>
<th>Unweighted ratings</th>
<th>Weights</th>
<th>Weighted ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRECIS</td>
<td>POPS</td>
<td>PRECIS</td>
</tr>
<tr>
<td>Published documentation</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Input string coding</td>
<td>9</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>Input string length</td>
<td>7</td>
<td>8</td>
<td>-1</td>
</tr>
<tr>
<td>Worksheets</td>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Input string detail</td>
<td>8</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Index string detail</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Index string eliminability</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Index string clarity</td>
<td>7</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

The summing of weighted ratings for each choice gave the overall values of 51 and 35 for PRECIS and POPS, respectively. In the light of the present study, it was decided to make the following changes:
<table>
<thead>
<tr>
<th>Features</th>
<th>Unweighted ratings</th>
<th>Weights</th>
<th>Weighted ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRECIS</td>
<td>DSIS</td>
<td>PRECIS</td>
</tr>
<tr>
<td>Published documentation</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Input string coding</td>
<td>9</td>
<td>6</td>
<td>-1</td>
</tr>
<tr>
<td>Input string length</td>
<td>6</td>
<td>9</td>
<td>-1</td>
</tr>
<tr>
<td>Worksheets</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Input string detail</td>
<td>8</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Index string detail</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Index string eliminability</td>
<td>8</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Index string clarity</td>
<td>7</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Overall collocation</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>No. of Cross references</td>
<td>8</td>
<td>4</td>
<td>-1</td>
</tr>
<tr>
<td>Null access points</td>
<td>0</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>Index bulk</td>
<td>6</td>
<td>9</td>
<td>-2</td>
</tr>
</tbody>
</table>

The overall ratings now stand as 44 and 27 for PRECIS and DSIS, respectively. However, it should be noted that, these ratings are based on mostly subjective observations of the present author, and certainly will vary in a different environment.
References


Bhattacharyya, G. (1982b). POPSI: A source language for organising and


heading 'Currency with coverage: Subject indexing proposals').


Classification Research Group (1955). The need for a faceted classification as the basis of all methods of information retrieval. Library Association Record, 57 (7): 262-268.


Coates, E. J. (1960). Subject catalogues: Headings and structure. London:


Devadason, F. J. (1986b). Personal communication.


Study Conference on Classification Research, Augsburg, FRG, June 28-
July 2, 1982 (pp. 86-101). Frankfurt, FRG: INDEKS Verlag; reprinted in
Chan, L. M., Richmond, P. A. and Svenonius, E. (Eds.), Theory of
subject analysis: A source book (pp. 392-402). Littleton, CO, USA:
Libraries Unlimited).

Press. (Rutgers series on systems for the intellectual organization of
information, ed. by S. Artandi, Vol. 4).


Social Science Information, 8: 17-42.


Ghose, A. and Dhawle, A. S. (1979). Between traditional classification and
coordinate indexing. Information Processing & Management, 15 (1): 27-
31.

Gilchrist, A. (1972a). Classification in the construction industry. Journal


Classification in the 1970s: A second look (pp. 51-80). Rev. ed.
London: Clive Bingley.

the classification of documents according to the UDC by adjusting its
symbols to the descriptors of the thesaurus.] (Translated title, LISA

Grolier, E. de (1962). A study of general categories applicable to
classification and coding in documentation.. Paris: Unesco.


service, Bangalore, India (Vol. 2, pp. 166-184). Bangalore, India: DRTC.


Rajan, T. N. (1976). The handling of compound terms in indexing systems.


with a Slant to Documentation, 5 (2): 97-133.


EXHIBITS

No.                                      Page No
I.  Manipulation Routines for DSIS Entries  336
II. Example of DSIS Input String and the Manipulation Matrix  337
III. Programming Problems Not Yet Fully Resolved in DSIS  338
IV.  DSIS Index String Generation Program  339
V.   Sample of PRECIS Index  347
VI.  Sample of DSIS Index  351
EXHIBIT - I
Manipulation routine for DSIS entries

Read next record
Store record number in n9$
Find next marker until nil left over
Between each marker pair is one term
For each term divide into codes (srm$ matrix) and alphanumeric data (trm$ matrix) - fmt% - linked by common counter variable.

Now using the code matrix:

Fill in the code matrix:
For each term which has $2 or $3 add 2 or 3 to Lead$
For each term which has $0 add 1 to Lead$ else 0
For each term which has $1 add 1 to cont(ext)$ else 0
For each new modifier of type 1 from < to > add to comp(lex)$
  for first term <
  for each element to last but one *
  for last element add >
else if not modifier comp$ = comp$ + SP

For each coordinate concept (i.e. Term code with extra zero)
  add 1 to coord$ else 0

Proceed to build context from 1 ... n
For each new element in matrix
  if cont$ = 0 goto next
  else if lead$ = 0 or 1 add directly to context
  else if 2 or 3 add to context and then
    add elements reading back to focus
Store context$

Now process Headings
For each element of Lead$ > 0
following routines called in context procedure
create Heading without codes
add to each heading which is not Modifier Type 1 reference
the context string plus n9$ (record number)
create sort key of minimum first 120 characters
write to entry file and sort file respectively
Example of input string and the manipulation matrix generated at one character per term

$9$ education$+860$$1$industries$+860$$1$banking$+860$$1$insurance$+860$$1$public utilities$+840$$personnel$+860$$1$clerical$+825$$performance$$+825$$1$effects of$+$2$flexible$$2$working hours$+$0$1$surveys$+$0$0$0$0$0$0$0$0$0$

n9$=ae0030$

Manipulation matrix

lead$: 011111310221 leads are $0,2,3$
cont$: 01111111111 contexts are $1$
comp$: <*> Type 1 modifiers
coord$: 001100000000 coordinate terms have 0 added to code
Programming problems not yet fully resolved in DSIS

1 Lack of a sort utility on the particular multi-user microcomputer has forced a compromise with the above stated minimum length of sort key (i.e. 120 characters). This length was found to be the minimum requirement necessary for a high proportion of the documents indexed by DSIS, and particularly for the micro-documents.

The technique which had to be adopted was to use a database manager index facility (Digital Research Access Manager using a B-tree index) with an initial sortkey length of 24 characters. (Note this was also used for the PRECIS indexes). Then when it came to the Print program, subsets of records, having one of these keys in common, were sent to a subsort routine in which the sortkeys were expanded to lengths of 120 characters).

This was successful for the basic form of DSIS with heading, context, but no upper-link specifiers. The length of sortkey for entries if these were to be included presented a difficult problem within the constraints of the system available before its enhancement in July 1988. By this date it was too late to be able to conduct proposed comparative tests in searching and retrieval.

2 The same problem was one of the root causes of failure to implement fully the Classaurus. Here we had to solve, it is thought for the first time, the problem of organisation of a multi-disciplinary classaurus. This raised wholly new problems of organisation if the classaurus was to be accessible on-line, and the programming skills necessary, particularly in the areas of file design and organisation, are still inadequate for this task. Without an operational classaurus, the generation of a complete index including equivalence references was impossible. Also the desirable enhancement of the input system could not be contemplated.

This enhancement is seen as fundamental to a practical working system of the future. Presently, thesaurus input follows the batch generation from the index strings as recommended by Devadason. Clearly, within any one discipline it is desirable that once a particular sub-tree has been entered into the classaurus, it should be possible to prepare subsequent index strings which incorporate the same tree without the repetition of all steps in that tree - it should in fact be possible (and desirable) to generate the upper links from the already existing classaurus records. One would envisage that when this desirable enhancement has been achieved, that the economical method for classaurus input would in the future be by an online facility, rather than in the batch mode.
EXHIBIT - IV

DSIS Index String Generation Program
rem reads a file of DSIS strings creates a context record for each
rem then for each subject heading writes that and context to entry
rem file - for references (modifier Type 1) writes reference to file

dim trm$(40),srm$(40)
on error goto 9000

DEF FNASC%(X%)=ASC(MID$(addr$,X%,1)) rem - reads packed address
rem
rem Random read from file on channel 'x'
rem returns record with quotes stripped off
rem

def fnum$(x,y)
ingTEGER x,y
string x$
xs=str$(x)$
x=len(xs)$
x$=left$("0000",y-x)$+x$
fnum$x=
s$
fend
def fnwrite(x$,y$)
string x$,y$
integer 19
x$x=x$+%$+$y$
19=len(x$)$
105 x$x=fnum$(19,3)$+x$
if len(x$)<125 then 200
print f3,filx6%; left$(x$,124) : filx6%=filx6%+1
x$x=right$(x$,len(x$)-124) : goto 200
200 -x$x=x$+string$(124-len(x$)," ")
print f3,filx6%; x$x : filx6%=filx6%+1
fend
def fnread$(x,y)
ingTEGER x,y,los
string a$
read f4,x$; line a$
if left$(a$,1)=chr$(34) then a$=right$(a$,len(a$)-1)
if right$(a$,1)=chr$(34) then a$=left$(a$,len(a$)-1)
10 fnread$x=a$
fend
rem using fnread$ to read individual record, reads in next string record
rem to length stored on first 3 characters

def fnstrg$(x)
ingTEGER x,n
18%=124
x$
if len(x$)<3 then 10
xn=x$n+1 : y$=fnread$(x,x$9%
18%=18%+124
x$x=x$+y$ : goto 10
20 fnstrg$x=mid$(x$4,n)
fend
rem
rem before next record zeros code matrix for last record

def fnzero(x)
ingTEGER x,p1
for p1=1 to x
srm$(p1)="" : trm$(p1)="
next p1
cont$="" : lead$="" : comp$="" : addr$="" : coord$=""
defdfcoord%(x$)
ingTEGER p,p1
string x$
if match("/",trm$(i%),1)>0 then fcoord%=0 : return
if match("E0",x$,1)<1 then fcoord%=0 : return
fcoord%=1
fend
rem Analyses codes - Case operators added to front of terms
rem Lead & context codes into A$ matrix
rem
rem If $0 then lead$=1 else =0 - Lead/non-Lead
rem If $1 then cont$=1 else =0 - Context/non-Context
rem if $3 then lead$=3 - Modifier difference
rem if trm$ starts with < then
rem comp$=< and trm$ starts with / and subsequent comp$=* till
rem trm=>$ when comp$=>
rem else comp$=space
rem if $* then lead$=0 : cont$=1 - Date
rem
def coder(code$)
    string code$, prk$
    integer, p, pl, p2

    p=match("$9", code$, 1)
    if p=1_ then code$="":
        lead=lead$+"0":
        cont$=cont$+"0":
        comp$=comp$+"":
        coord$=coord$+"0": goto 200
    p=match("$", code$, 1)
    if p<1 then prk$="": goto 5

    prk$=right$(code$, len(code$)-(p-1)): code$=left$(code$, p-1)

    5 if left$(trm$(i%), 1)="<" then kind2%=1 : comp$=comp$+"<":
        trm$(i%)=right$(trm$(i%), len(trm$(i%))-1):
        goto 10
    if right$(trm$(i%), 1)=">" then kind2%=0 : comp$=comp$+">":
        trm$(i%)=left$(trm$(i%), len(trm$(i%))-1): goto 10
    if kind2%>0 then comp$=comp$+"*: goto 10
        comp$=comp$+"

    10 trm$(i%)="+code$+"+trm$(i%)"
    if fcoord%(code$) then coord$=coord$+"1": goto 15
        coord$=coord$+"0"

    15 if match("$3", prk$, 1)>0 and
        match("$0", prk$, 1)<1 then lead$=lead$+"4": goto 20
    if match("$3", prk$, 1)>0 then lead$=lead$+"3": goto 20
    if match("$2", prk$, 1)>0 then lead$=lead$+"2": goto 20
    if match("$0", prk$, 1)>0 then lead$=lead$+"1":
        else lead$=lead$+"0"

    20 if left$(code$, 1)="2" then prk$="$1"
    if left$(code$, 1)="0" and len(prk$)<2 then prk$="$1"
    if match("$1", prk$, 1)>0 or
        match(mid$(comp$, i%, 1), "<>", 1)>0 then cont$=cont$+"1":
        else cont$=cont$+"0"

    200 addr$=addr$+chr$(1%)
    srm$(i%)=code$

fend
rem Analyse POPSI string
rem
def $pog$c$y$k$, x$
    integer p, pl, x3, los
    pl=1 : x3=1 : los=len(rec$)

10 p=match("+", rec$, p1)
    if p<1 then 300
    prk$=mid$(rec$, p1.p-1)

    p3=len(prk$)
    pl=match("$", prk$, 1) rem Get rid of terms which are dates
    if pl> then pl=pl+2 : goto 50
    pl=match("1", prk$, 1) rem Find first letter of term
    20 if pl<1 then print "incomplete term **"; prk$ : goto 200
    if match(mid$(prk$, pl-1,1), "(<", 1)>0 then pl=pl-1
rem To find beginning of terms which start with numerals
for pl=p1 to 1 step -1
    if match(mid$(prk$,p1,1),".$",1)>0 then p1=p1+2 : goto 50
next pl
goto 20

50 trm$(x3)=right$(prk$,p3-p1+1) : srm$(x3)=left$(prk$,p1-1) :
   goto 200

200 if right$(trm$(x3),1)=" " then \
   trm$(x3)=left$(trm$(x3),len(trm$(x3))-1) : goto 200
210 if left$(trm$(x3),1)=" " then \
   trm$(x3)=right$(trm$(x3),len(trm$(x3))-1) : goto 210
x3=x3+1 : pl=pl+1
   if pl<los then 10
fend

rem for $3 differences creates term read from R to L to focus
rem
def diff$(p)
   integer p,pl
   string x$,y$
   diff$="" : x$=""
   p1=match("!",trm$(p),1)
   if p1<1 then 50
   x$=left$(trm$(p),p1-1)
   if match(".$0",x$,1)>0 then blckflg%=1 else blckflg%=0
   50 y$=trm$(p)+" "
   p1=match("!",y$,1)
   y$=right$(y$,len(y$)-(p1-1))
   x$=x$+y$
   60 p=p-1
   if p<1 then 100
   if blckflg<1 then 80
   if mid$(lead$,p,1)="3" and mid$(coord$,p,1)="1" then 60
   80 if match(mid$(lead$,p,1),"234",1)>0 then 50
   90 y$=trm$(p)
   p1=match("!",y$,1)
   y$=right$(y$,len(y$)-(p1-1))
   x$=x$+y$
   100 diff$=x$
fend

rem to remove codes from headings and optionally contexts
rem
def stripcod$(x$)
   string x$,y$
   y$€™s p,p1
   for p=1 to len(x$)
      if match(mid$(x$,p,1),".$0123456789/-",1)>0 then 10
      y$=y$+mid$(x$,p,1)
   10 next p
   20 if left$(y$,1)=" " then y$=right$(y$,len(y$)-1) : goto 20
   30 p=match(" ",y$,1)
   if p>0 then y$=left$(y$,p-1)+right$(y$,len(y$)-p) : goto 30
   stripcod$=y$
fend

def alpha$(x$)
   string x$
   integer p,p1
   p=match("!",x$,1)
   if p<2 then 10
      if mid$(x$,p-1,1)="(" then p=p-1
      x$=right$(x$,len(x$)-(p-1))
   10 p=match(".$0",x$,1)
      if p<1 then 25
   x$=left$(x$,p-2)+right$(x$,len(x$)-p-3) : goto 10
25
p = match(" ", x$, 1)
if p>0 then x$=left$(x$,p-1)+right$(x$,len(x$)-p) : goto 40
x$=stripcod$(x$)
alpha$=x$
fend
rem
to create coordinate block of terms
rem
def fblock$(x)$
integer x, p, pl, diff
string x$, y$
x$="" y$=srms$(x$) rem - store term operator for block
10 p = match("O", coord$, x+1)-1
if p<0 then p = len(coord$)
on val(mid$(lead$, x, 1))+1 goto 120, 15, 15, 40, 40
diff = 0
20 for x=x to p-1
   e% = fnasc%(x)
   if srm$(x+1)<>y$ then 25
   if diff > 0 then x$=x$+" +trm$(e%)
24 next x
25 x$=x$+" (and) "
   if diff > 0 then x$=x$+diff$(x) : goto 30
   x$=x$+trm$(fnasc%(x))
30 i% = x
   fblock$=x$ : return
40 rem Block is composed of differenced terms
rem set diff flag to yes
diff = 1 : goto 20
120 x = x+1
   if mid$(coord$, x, 1)="1" then 10
   i% = x
fend
rem
creates Type 1 modifier main heading and references
rem
def ref$(p)$
   anag#$x$, p$, p2, p3
   x$="" : y$=""
rem
rem p=pos of Lead term
rem pl=pos to begin ->
rem p2=pos of end ->
   for pl = p to 2 step -1
      if mid$(comp$, pl, 1)="<" then 10
next pl
10 p2 = match(">", comp$, p, 1+1)
   if p = p1 then 50
   if mid$(lead$, p, 1)="3" then x$=alpha$(diff$(p))+" " : goto 50
   p3 = p+1
   x$=alpha$(trm$(fnasc%(p)))+" "
   if mid$(lead$, p3, 1)="2" then x$=x$+alpha$(trm$(fnasc%(p3))) : 
   p3 = p3+1 : goto 12
goto 50
40 if len(x$)<1 then x$=y$ else x$=x$+" / "+y$
   ref$=x$ : return
50 if mid$(lead$, p, 1)="2" then 55
   if p = p2 and mid$(lead$, p2, 1)="<" then 55
   if mid$(lead$, p, 1)="2" and match("0", lead$, p, 1) then 1 then p2 = p-1
55 for p1 = p to p2
   if match("/£", trm$(p1, 1)) then 60
   if mid$(coord$, p1, 1)="1" then y$=y$+\fblock$(p1) : p1 = i% : goto 60
   if len(y$)>0 then y$=y$+" "
   if mid$(lead$, p1, 1)="3" then y$=y$+alpha$(diff$(p1)) : goto 60
   if p1 = j% then y$=y$+@ " : goto 60
\[ y = y + \alpha (trm(fnasc(pl))) \]

\[ 344 \]

\[ 60 \text{ next } pl \]

\[ \text{goto 40} \]

\[ \text{fend} \]

\[ \text{def complex$(x)$} \]

\[ \text{integer } x, y \]

\[ \text{string } x$, y$ \]

\[ \text{if mid$(coord$, x, 1)="1" then } x$=alpha$(fblock$(x)) +" " : y=j% : \text{goto 10} \]

\[ p1=\text{match("!", trm$(x)$, 1)} \]

\[ \text{if } p1<1 \text{ then } p1=1 \]

\[ \text{if mid$(trm$(x)$, p1-1, 1)="(" then } p1=p1-1 \]

\[ y$=\text{right$(trm$(x)$, len(trm$(x)$)-p1+1)} \]

\[ x$=x$+y$+" " \]

\[ y=x \]

\[ 10 \text{ for } y=y+1 \text{ to len(comp$)} \]

\[ \text{if match(mid$(comp$, y, 1), ",", 1)>0 \text{ then 50}} \]

\[ \text{if match("/", trm$(y)$, 1)>0 \text{ then 40}} \]

\[ p1=\text{match("!", trm$(y)$, 1)} \]

\[ \text{if } p1<1 \text{ then } p1=2 \]

\[ \text{if mid$(trm$(y)$, p1-1, 1)="(" then } p1=p1-1 \]

\[ y$=\text{right$(trm$(y)$, len(trm$(y)$)-p1+1)} \]

\[ x$=x$+y$+" " \]

\[ 40 \text{ next } y \]

\[ 50 \text{ x$=\text{stripcod}(x$)} \]

\[ \text{complex$=x$ : x$=""} \]

\[ \text{fend} \]

\[ \text{rem} \]

\[ \text{rem} \text{ to print entries and strings on VDU} \]

\[ \text{rem} \]

\[ \text{def fprt(x$, y$, strg$, mark$)} \]

\[ \text{string } strg$, mark$ \]

\[ \text{print } x$; »s56\text{len(strg$)} \]

\[ \text{pl%}=1 \]

\[ 100 \text{ if } h%=\text{los%} \text{ then 400} \]

\[ \text{for } h%=h% \text{ to } 1 \text{ step } -1 \]

\[ \text{if match(mid$(strg$, h%, 1), mark$, 1)>0 \text{ then 150}} \]

\[ \text{next } h% \]

\[ \text{goto 400} \]

\[ 150 \text{ print mid$(strg$, pl%, h%-(pl%)-1))} \]

\[ \text{print } y$; \]

\[ \text{pl%=h%+1 : h%=pl%+54 : } \text{goto 100} \]

\[ 400 \text{ print right$(strg$, los%-(pl%)-1))} \]

\[ \text{fend} \]

\[ \text{rem} \]

\[ \text{rem} \text{ Constructs context of entry} \]

\[ \text{rem} \]

\[ \text{rem} \text{ ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::} \]

\[ \text{rem} \text{ ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::} \]

\[ \text{rem} \text{ ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::} \]

\[ \text{rem} \text{ START HERE} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{rem} \]

\[ \text{input "Filename "; line } h$ \]

\[ \text{if right$(h$, 4)<>". tra" then } h$=h$+". tra" \]

\[ \text{open } h$ \text{ recl 128 as 1} \]

\[ \text{read } l1,1; \text{ line rec$} \]

\[ x8%=\text{val(mid$(rec$, 2, 4))}-1 \]

\[ \text{input "Start at "}; \text{ x9%} \]

\[ \text{input "Finish at "}; \text{ x5%} \]

\[ \text{input "Entries file "; line out$} \]

\[ \text{create out$ recl 128 as 3} \]

\[ \text{filx6%=1} \]

\[ \text{rem} \]

\[ 10 \text{ if end } l1 \text{ then 9999} \]
rec$=fnstrg$(1)
x9%*x9%+1
  p%=match("%", rec$, len(rec$)-6)
  if p%<1 then 20
  if x9%<2 or x9%>x8% then 9990
  if p%<1 then 20
  n9$=right$(rec$, len(rec$)-p%)
  rec$=left$(rec$, p%-1)
  print n9$
call fprt("", ", rec$, ". 0248$"
if constat% then input line mm$ : if len(mm$)<1 then goto 9990
  20 gosub 100
if x9%>x5% then 9990
goto 10
100 fmat%=fopens$(rec$)
  kind2%=0
  for i%=1 to fmat%
    print srm$(i%), trm$(i%)
    call coder(srm$(i%))
  next i%
  print lead$
  print cont$
  print comp$; "**"
  print coord$; "**"
1000 context$="": x%=0
  if len(mm$)<1 then goto 9990
  for i%=1 to fmat%
    print srm$(i%), trm$(i%)
    call coder(srm$(i%))
  next i%
  print lead$
  print cont$
  print comp$; "**"
  print coord$; "**"
1050 if mid$(coord$, i%, 1)="1" then context$=context$+" +fblock$(i%)
  goto 990
rem if mid$(comp$, i%, 1)="0" and match(mid$(lead$, i%, 1), "34", 1)<1 then 1200
1100 if match(mid$(lead$, i%, 1), "34", 1)>0 \
  then context$=context$+diff$(i%)
  goto 990
rem if len(srm$(i%))>1 then trm$(i%)=superord$(i%)
  context$=context$+trm$(i%) : goto 990
1990 next i%
rem tidies up context
rem
1990.1 if left$(context$, 1)=" " then \ 
  context$=right$(context$, len(context$)-1) : goto 1990.1
1991 if match(".£", left$(context$, 3), 1)>0 then \ 
  context$=left$(context$, 1)+right$(context$, len(context$)-3) : \ 
  goto 1991
1992 p%=match("/", context$, 1)
  if p%<1 then 1993
  context$=left$(context$, p%)+right$(context$, len(context$)-p%-1) : \ 
  goto 1992
1993 p%=match(" ", context$, 1)
  if p%<1 then 1994
  context$=left$(context$, p%-1)+right$(context$, len(context$)-p%)
  goto 1993
1994 p%=match(" £0", context$, 1)
  if p%>0 then \ 
    context$=left$(context$, p%+1)+right$(context$, len(context$)-p%-2) : \ 
    goto 1994
  for p%=1 to 3
1996 p%=match(mid$("842", p1%, 1)+"0", context$, 1)
  if p%<1 then 1998
  if mid$(context$, p%-1, 1)=" " then 1997
  if mid$(context$, p%-1, 1)<chr$(47) and mid$(context$, p%-1, 1)<chr$(58) then 1997
  context$=left$(context$, p%)+right$(context$, len(context$)-p%-1) : goto 1996
1998 next p1%
rem
rem creates successive leads for lead matrix string 1,2, or 3
for j%=1 to 1%
  lead%=val(mid$(lead$, j%, 1))+1
  if mid$(coord$, j%, 1)<"0" and mid$(comp$, j%, 1)<" " then 2200
    on lead% goto 2990, 2050, 2200, 2400, 2400
rem 2050
  if mid$(comp$, j%, 1)<" " then 1$=complex$(j%): goto 2800
  1$=alpha$(trm$(fnasc%(j%)))
  goto 2800
2100 rem Kind 1 - main entry
  1$=ref$(j%): goto 2800
rem 2200 rem Kind 1 - reference
rem 2200 rem
  if mid$(comp$, j%, 1)<" " then 2100
  1$=ref$(j%)
2210 p%=match("/", 1$, 1)
  if p%>0 then 1$=left$(1$, p%)+right$(1$, len$(1$)-p%)-1): goto 2210
2220 p%=match(" ", 1$, 1)
  if p%>0 then 1$=left$(1$, p%-1)+right$(1$, len$(1$)-p%): goto 2220
  call fprt(" ", " ", 1$, ", ", " ")
  1$=1$+"*": goto 2900
2400 if mid$(comp$, j%, 1)<" " then 2100
  1$=diff$(j%): 1$=alpha$(1$)
2800 1$=1$+"*"+context$
  call fprt(" ", " ", 1$, ", ", " "")
2900 call fnwrite(1$, n9$)
2990 1$=""
  next j%
3000 call fnzero(1%)
return
9000 i$=err: i%=errl
  print "error "; i$; " at "; i%
9990 close 1, 3
EXHIBIT - V: Sample of PRECIS Index

Students
See also
Adult students
Postgraduates
Students
Academic course students. Learning patterns compared with learning patterns of vocational education students == AE0008
Attitudes to learning patterns & teaching methods == AE0033
Attitudes to shortcomings of American education == AE0089
Primary health care students. Problem solving skills. Assessment. Applications of computer systems == AE0001
Students. Adult basic educational institutions & adult secondary educational institutions
Recruitment & retaining. Role of professional tutors == AE0063
Students. Chemeketa Community College
Vocational education. Two-year courses: Communication skills == AE0002
Students. Further education institutions
Learning == AE0135
Students. Further education institutions. Great Britain
Academic achievement. Assessment - For teaching == AE0197
Students. Primary schools. England
Admission. Policies of local education authorities.
Research == AE0039
Students. Primary schools. Great Britain
Academic achievement. Effects of ability, family background, sex & sociocultural factors. Research == AE0040
Students. Vocational education institutions
Learning patterns compared with learning patterns of academic course students == AE0008
Recruitment. Methodology == AE0023
Students, 16-19 years. Colleges of further education. Scotland
Education == AE0121
Students, 16-19 years. Further education institutions. England
One-year courses - Reports == AE0146
Students, grades 11 & 12
Academic education & vocational education. Projects: 'Principles of Technology' == AE0031
Academic education & vocational education. Use of laboratory equipment & video equipment == AE0031
Study techniques. Adult education == AE0180
Supply. Teachers. Industrial education institutions & trade education institutions
- Surveys == AE0088
Task groups
Decision-making behaviour == AE0035
Tax-credit system. Tuition fees. Schools
Economic aspects == AE0054
Teacher tutors See Professional tutors
Teachers
See also
Faculties
Part-time teachers
Tutors
Teachers
Attitudes to learning patterns & teaching methods == AE0033
Teachers
Home economics teachers. Attitudes to role of unemployment in family problems == AE0073
Industrial arts teachers. Professional education.
Programmes. Enrollment. Decision making factors == AE0092
Teaching skills == AE0004
Teachers. Industrial education institutions & trade education institutions
Supply & demand - Surveys == AE0088
Teachers. Personnel. Correctional education
Selection. Criteria == AE0051
Teachers. Personnel. Schools
In-service training. Programmes. Role in development of local careers guidance services == AE0069
Teachers. Primary schools & secondary schools. Mississippi. United States
Professional education. Curriculum. Development. Role of professional tutors == AE0068
Teaching
See also
Learning
Tutoring
Teaching
Great Britain. Adult education - For teaching == AE0188
Great Britain. Further education institutions. Students.
Academic achievement. Assessment - For teaching == AE0197
Use of visual aids: Transparencies - Manuals == AE0003
Teaching. Adult education
Theories == AE0133
Teaching aids
See also
Visual aids
Teaching aids. Adult literacy education & adult numeracy education. England
Worksheets == AE0200
Teaching aids. Competency-based education. Trainees Interactive computer video systems == AE0034
Teaching aids. Self-teaching. Further education. Great Britain
Provision == AE0184
Teaching methods
See also
Classroom techniques
Distance study
Student-centred teaching methods
Teaching methods
Attitudes of students & teachers == AE0033
Teaching methods. Adult education == AE0180
Teaching methods. Further education institutions Applications of computer systems == AE0112
Teaching methods. Reading. English as a second language Classroom techniques == AE0071
Teaching skills. Teachers == AE0004
Teaching support services. Further education institutions.

Great Britain
Management == AE0182

Technical and Vocational Educational Instruction
See Technical education, Vocational education

Technical education
related to community development == AE0009

Technical education. Arkansas. United States
Public technical education. Organisation structure == AE0027

Technical education. Community colleges
Concepts == AE0026

Technical schools. United States
Regional vocational technical schools == AE0005

Technician Education Council courses. London & Home Counties. England
- Directories - Serials == AE0122

Technology. Allied health profession & medical profession
Impacts on historical evolution == AE0074

Technology. Curriculum subjects. Continuing education
Campus-based industrial technology & military-based
industrial technology. Attitudes of graduates - Comparative studies == AE0084
Campus-based industrial technology & military-based
industrial technology. Attitudes of graduates - Comparative studies == AE0084

Technology. Curriculum subjects. Educational institutions
Industrial technology. Change == AE0060

Tertiary education
See also
Higher education

Testing See Assessment
Tests
See also
Objective tests

Educational institutions
Concepts: Private enterprise - Surveys == AE0055
Theories. Curriculum. Adult education == AE0155
Theories. Teaching. Adult education == AE0133

Trade education institutions
Teachers. Supply & demand - Surveys == AE0088

Trade unions
Role in careers guidance - Surveys == AE0067
Role in industrial health & safety in chemical industries
== AE0044

Trade unions. Great Britain
Attitudes to part-time employment == AE0065

Traditional occupations
Traditional female occupations. Women. Socialisation - Review of literature == AE0064

Trainees
Competency-based education. Teaching aids: Interactive computer video systems == AE0034

Trainees. Great Britain
Youth Opportunities Programme trainees. Social education - For management == AE0179

Trainers
Role in industrial health & safety in chemical industries
== AE0044
Training

See also
In-service training

   Educational institutions
   Programmes. Applications of computer systems == AE0059

Training. Personnel. Further education institutions. Great Britain
   Curriculum subjects: Computer systems - Conference proceedings == AE0114

Training programmes
   Role in industrial health & safety in chemical industries == AE0044

   Monitoring by Office of Personnel Management == AE0010

Transparencies. Visual aids
   Use in teaching - Manuals == AE0003

Tuition fees. Schools
   Tax-credit system. Economic aspects == AE0054

Tutoring. Further education == AE0134

TVEI See Technical education, Vocational education

Two-year courses. Vocational education. Students.
   Chemeketa Community College
   Communication skills == AE0002

Typing
See also
   Keyboarding

Uncertainty. Employment or retirement. Old personnel == AE0058

Unemployed adults. Great Britain
   Further education == AE0183

Unemployed persons. Great Britain
   Unemployed young persons. Vocational preparation.
      Programmes: Youth Opportunities Programme. Work Introduction Courses == AE0161

Unemployment
   Role in family problems. Attitudes of consumers and home economics teachers == AE0073

Unemployment. South Yorkshire (Metropolitan County).
   Yorkshire. England
   Implications for further education == AE0128

Universities
   Counselling centres. Anonymous clients. Satisfaction compared with satisfaction of self-identified clients - Surveys == AE0046
   Counselling centres. Self-identified clients. Satisfaction compared with satisfaction of anonymous clients - Surveys == AE0046

Unskilled personnel
   Attitudes to retirement. Implications for pre-retirement education == AE0090

Video equipment
   Applications in competency-based education & individualised instruction == AE0032
   Use in academic education & vocational education.
      Students, grades 11 & 12 == AE0031

Video systems. Teaching aids. Competency-based education.
   Trainees
   Interactive computer video systems == AE0034
Students
Education 8 Students 8.1 Learning patterns 8.1.5 (attitudes of) Students 8.1.5 (and) Teachers == 0033
Education 8 Students 8.6 Academic course students 8.2 Learning patterns 8.2.5 (compared to) Vocational students' learning patterns == 0008
Education 8 Students 8.6 Health care students 8.6 Primary health care students 8.2 Skills 8.2.6 Problem solving skills 8.2.1 Assessment 8.1.5 (using) Computer systems == 0001
Education 8 Students 8.6 Health students 8.6 Allied health students 8.1 Teaching 8.1.4 Special subjects 8.1.6 Ethics 8.1.6 Biomedical ethics 8.1.4 Skills 8.1.6 Decision making skills 0 Case studies == 0075
Education 9.6 Adult education 8 Students 8.6 External students 8.6 Adult external students 8.1 Examinations 8.1.6 G.C.E. examinations 4 Great Britain 0 Inquiry reports == 0173
Education 9.6 Basic education (and) 9.6 Secondary education 9.6 Adult secondary basic education 8 Students 8.1 Recruitment (and) 8.1 Retaining 8.1.5 (role of) Professional tutors == 0063
Education 9.6 Basic education (and) 9.6 Secondary education 9.6 Adult secondary basic education 8 Students 8.1 Recruitment (and) 8.1 Retaining 8.1.5 (role of) Professional tutors == 0064
Education 9.6 Development education 8 Students 8.1 Learning strategies 8.1.1 Improvement 8.1.5 (role of) Oxfam America 4 Developing countries == 0041
Education 9.6 Higher education 8 Students 8.6 Adult students 8.6 Middle-aged adult students 8.1 Development == 0049
Education 9.6 Vocational education 8 Students 8.2 Learning patterns 8.2.5 (compared to) Academic course students' learning patterns == 0008

Students (in) Further education institutions
Education 9.6 Further education 8 Students 8.5 (in) Further education institutions 8.1 Learning == 0135
Education 9.6 Further education 8 Students 8.5 (in) Further education institutions 8.1 Academic achievement 8.1.1 Assessment 4 Great Britain 0 For teachers == 0197

Students (in) Vocational educational institutions
Education 9.6 Vocational education 8 Students 8.5 (in) Vocational educational institutions 8.1 Recruitment 8.1.4 Methodology == 0023

Students / Learning patterns (attitudes of) Students (and) Teachers

Students / Shortcomings (attitudes of) Faculties (and) Personnel (and)

Students / Teaching methods (attitudes of) Students (and) Teachers

Students / Twoterm (in) Communication skills (for) Students (in) Community colleges

Students' learning patterns / Learning patterns (compared to) Academic course students' learning patterns
Students' learning patterns / Learning patterns (compared to) Vocational students' learning patterns

Students, grades
Education 9.6 Academic education (and) 9.6 Vocational education 9.5 (using) Laboratory equipments 9.5 (and) Video equipments 8 Students, grades 11-12 == 0031

Students, years (in) Colleges of further education
Education 9.6 Further education 8 Students, 16-19 Years 8.5 (in) Colleges of further education 4 Great Britain 4.4 Scotland == 0121

Students, years / Oneyear courses (for) Students, years (in) Further education institutions

Study techniques
Education 9.6 Adult education 9.4 Study techniques (and) 9.4 Teaching methods == 0180

Supply
Education 8 Education institutions 8.6 Industrial education institutions 8.6 Trade education institutions 8.6 Industrial education institutions and trade education institutions 8.4 Teachers 8.1 Demand 8.1 Supply 8.1 Demand and supply 0 Surveys == 0088

Task groups
Education 8 Groups 8.6 Task groups 8.1 Behaviour 8.1.6 Decision-making behaviour == 0035

Tax credits
Education 8 Schools 8.1 Tax credits 8.1.6 Tuition tax credits 8.1.6 Economic aspects == 0054

Teacher tutors See Professional tutors

Teachers
Education 8 Education institutions 8.6 Industrial education institutions 8.6 Trade education institutions 8.6 Industrial education institutions and trade education institutions 8.4 Teachers 8.1 Demand 8.1 Supply 8.1 Demand and supply 0 Surveys == 0088
Education 8 Teachers 8.1 Teaching methods 8.1.5 (attitudes of) Students 8.1.5 (and) Teachers == 0033
Education 8 Teachers 8.2 Skills 8.2.6 Teaching skills == 0004
Education 8 Teachers 8.6 Industrial arts teachers 8.1 Professional education 8.1.6 Programmes 8.1.1 Enrollment 8.1.6 Decision making factors == 0092
Education 9.6 Correctional education 8 Teachers 8.1 Selection 8.1.4 Criteria == 0051
Education 9.6 Development education 8 Teachers 8.1 Teaching methods 8.1.1 Improvement 8.1.5 (role of) Oxfam America 3 Developing countries == 0041
Education 9.6 Professional education 8 Teachers 8.6 Part-time teachers 8.5 (in) Adult education institutions 8.5 (and) Further education institutions 4 Great Britain 4.4 England == 0152

Teachers / Family problems (role of) Unemployment (attitudes of) Consumers (and) Home economics teachers
Teachers / Learning patterns (attitudes of) Students (and)
Teachers / Professional education (for) Primary school teachers (and) Secondary school teachers
Teachers / Teaching methods (attitudes of) Students (and)

Teaching
Education 8 Students 8.6 Health students 8.6 Allied health students 8.1 Teaching 8.1.4 Special subjects 8.1.6 Ethics 8.1.6 Biomedical ethics 8.1.4 Skills 8.1.6 Decision making skills 0 Case studies == 0075
Education 9.6 Adult education 9.4 Teaching 9.4 Theories == 0133

Teaching (using) Visual aids
Education 9.4 Teaching 9.5 (using) Visual aids 9.6 Transparencies 0 Manuals == 0003

Teaching aids
Education 9.6 Further education 9.4 Self-teaching 8 Teaching aids 8.1 Provision 4 Great Britain == 0184
Education 9.6 Literacy education 9.6 Adult literacy education 8 Teaching aids 8.6 Worksheets 4 England == 0200

Teaching methods
Education 9.6 Adult education 9.4 Study techniques (and) 9.4 Teaching methods == 0180
Education 9.6 Development education 8 Teachers 8.1 Teaching methods 8.1.1 Improvement 8.1.5 (role of) Oxfam America 3 Developing countries == 0041
Education 9.6 Language education 8 English as a second language 8.2 Reading skills 8.2.1 Teaching methods 8.1.6 Classroom techniques == 0071
Education 9.6 Vocational education 8 Educational institutions 8.6 Training institutions 8.6 Vocational training institutions 8.6 Saint Petersburg Vocational Technical Institute 8.1 Teaching methods 8.1.6 Competency-based instruction 4 United States 4.4 Florida == 0006

Teaching methods (attitudes of) Students (and) Teachers
Education 8 Teachers 8.1 Teaching methods 8.1.5 (attitudes of) Students 8.1.5 (and) Teachers == 0033

Teaching methods (using) Computer systems
Education 9.6 Further education 8 Education institutions 8.6 Further education institutions 8.1 Teaching methods 8.1.5 (using) Computer systems == 0112

Teaching practice
Adult education 8 Schools 8.1 Curriculum assessment 8.1 Organisation structure 8.1 Staffing (and) 8.1 Teaching practice 4 Great Britain == 0036

Teaching skills
Education 8 Teachers 8.2 Skills 8.2.6 Teaching skills == 0004

Teaching support services
Education 9.6 Further education 8 Education institutions 8.6 Further education institutions 8.4 Teaching support services 8.1 Management 4 Great Britain == 0182
Technical and Vocational Educational Instruction

See Technical education, Vocational education

Technical education

Education 9.6 Industrial arts education (and) 9.6 Technical education 9.1 Change == 0060
Education 9.6 Technical education (and) 9.6 Vocational education 8 Educational institutions 8.6 Colleges 8.6 Community colleges == 0026
Education 9.6 Technical education 9.6 Vocational technical education 8 Schools 8.6 Technical schools 9.6 Vocational technical schools 9.6 Regional vocational technical schools 4 United States == 0005
Education 9.6 Technical education 9.6 Vocational technical education 9.6 Public vocational technical education 9.1 Evaluation 4 United States 4.4 Arkansas == 0027

Technical education (related to) Community development
Education 9.6 Technical education 9.5 (related to) Community development == 0009

Technical schools

Education 9.6 Technical education 9.6 Vocational technical education 8 Schools 8.6 Technical schools 9.6 Vocational technical schools 9.6 Regional vocational technical schools 4 United States == 0005
Technician Education Council courses
Education 9.6 Further education 8 Courses 8.5 (in) Further education institutions 8.6 Business Education Council courses 8.6 Ordinary National Diploma courses (and) 8.6 Technician Education Council courses 4 Great Britain 4.4 England 4.4 London and Home Counties 0 Directories 0 Serials == 0122

Technology

Education 8.1 Continuing education 8.1.4 Curriculum subjects 8.1.6 Technology 8.1.6 Industrial technology 8.1.6 Campus-based industrial technology 8.1.6 Military-based industrial technology 8.1.6 Campus-based industrial technology and military-based industrial technology 8.1.5 (attitudes of) Graduates 0 Comparative studies == 0084

Technology / Historical evolution (impacts of) Specialization (and)

Testing See Assessment

Textbooks

Education 8 Curriculum subjects 8.5 (in) Educational institutions 8.6 Economics (and) 8.6 History 8.4 Textbooks 8.4 Concepts 8.6 Free enterprise systems 0 Surveys == 0055

Theories

Education 9.6 Adult education 9.4 Curriculum 9.4 Theories == 0155
Education 9.6 Adult education 9.4 Teaching 9.4 Theories == 0133

Trade education institutions

Education 8 Education institutions 8.6 Industrial education institutions 8.6 Trade education institutions 8.6 Industrial education institutions and trade education institutions 8.4 Teachers 8.1 Demand 8.1 Supply 8.1 Demand and supply 0 Surveys == 0088
Trade unions / Career education (role of) Business firms
(and) Industries (and)
Trade unions / Occupational health & safety (role of) Hazard
Communication Rule (and) Trade unions (and) Trainers (and)
Training programmes
Trade unions / Parttime employment (attitudes of)
Trainees
Education 9.6 Competency-based education 9.5 (using)
Computer interactive video systems 8 Trainees == 0034
Education 9.6 Social education 8 Trainees 8.6 Youth
Opportunities Programme trainees 4 Great Britain 0 For
managers == 0179
Trainers / Occupational health & safety (role of) Hazard
Communication Rule (and) Trade unions (and) Trainers (and)
Training programmes
Training
Education 8 Personnel 8.1 Training 8.1.6 In-service
training 8.1.5 (role of) Conferences == 0014
Education 9.6 Further education 8 Personnel 8.5 (in)
Further education institutions 8.1 Training 8.1.4 Curriculum
subjects 8.1.6 Computer systems 4 Great Britain 0 Conference
proceedings == 0114
Education 9.6 Marketing & distributive 9.1 Training 9.1.6
Programmes 9.1.5 (using) Computer systems == 0059
Training institutions
Education 9.6 Vocational education 8 Educational
institutions 8.6 Training institutions 8.6 Vocational
training institutions 8.6 Saint Petersburg Vocational
Technical Institute 8.1 Teaching methods 8.1.6
Competency-based instruction 4 United States 4.4 Florida ==
0006
Training programmes / Occupational health & safety (role of)
Hazard Communication Rule (and) Trade unions (and) Trainers
(and)
Training schemes (for) Personnel (in) Federal Government
Education 9.6 Inservice education 8 Training schemes 8.5
(for) Personnel 8.5 (in) Federal Government 8.1 Monitoring
8.1.5 (by) Office of Personnel Management 4 United States ==
0010
Transparencies
Education 9.4 Teaching 9.5 (using) Visual aids 9.6
Transparencies 0 Manuals == 0003
Tuition tax credits
Education 8 Schools 8.1 Tax credits 8.1.6 Tuition tax
credits 8.1.6 Economic aspects == 0054
Tutoring
Education 9.6 Further education 9.1 Tutoring == 0134
Tutors / 8.10 Recruitment (and) 8.10 Retaining (role of)
Professional tutors
Tutors / Development (role of) Professional tutors
TVEI See Technical education, Vocational education
Two-term (in) Communication skills (for) Students (in)
Community colleges
Education 9.6 Vocational education 8 Courses 8.6 Two-term
courses 8.5 (in) Communication skills 8.5 (for) Students 8.5
(in) Community colleges 8.6 Chemeketa Community College 4
United States 4.4 Oregon == 0002