LINGUISTIC ANALYSIS AND ITS APPLICATION IN INFORMATION RETRIEVAL TOOLS

THESIS SUBMITTED
FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

R. C. PANDEY

DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE
NORTH - EASTERN HILL UNIVERSITY
SHILLONG
1996

पूर्वोत्तर पर्वतीय विश्वविद्यालय
मयूरभंज परिसर, शिलांग - 793014 (मेघालय)
North-Eastern Hill University
Mayurbhanj Complex, Shillong - 793014 (Meghalaya)

Phone :
Grams : NEHU

May, 1996

Prof. J.C. Binwal
Professor & Head
Department of Library
and Information Science

CERTIFICATE

This is to certify that the thesis entitled **LINGUISTIC ANALYSIS AND ITS APPLICATION IN INFORMATION RETRIEVAL TOOLS**, submitted for the award of the Degree of **DOCTOR OF PHILOSOPHY** in the Library and Information Science, is a faithful record of the bonafide research work carried out by **SHRI R.C. PANDEY** in the Department of Library and Information Science, North-eastern Hill University, Shillong - 79 30 14, under my guidance and supervision.

To the best of my knowledge, the work is in original and no part of the thesis has been submitted by him to this university for any other degree or diploma, and is worthy of consideration for the award of Doctor of Philosophy Degree in Library and Information Science.

JCBinwal

Date : 10th May, 1996.

Prof. J.C. Binwal

Place: Shillong.

Supervisor and Head
(Seal)
Head
Dept. of Library and
Information Science

ACKNOWLEDGEMENTS

It is indeed a matter of great pleasure and satisfaction to record the constant guidance, encouragement and assistance received from a variety of sources during the whole process of research starting from finalization of the topic to submission of the thesis for evaluation.

At the very first instance, I would like to express my deep sense of gratitude to Professor J.C. Binwal, my supervisor for his valuable guidance and inspiration with intellectual stimuli to complete the research work in time, when I had any problem, his helpful guidance put me back on track. I am indebted to him for his kindness and patience to listen to me during the course of research.

I would like to express my sincere thanks to the authorities of Central Institute of English & Foreign Languages, Hyderabad, and Shillong, for providing me the necessary facilities to pursue the research work. In this regard, I am also grateful to Professor J.C. Mahanti, Director, and Dr. R.K. Jayaraman, Former Officer-

in-Charge of the Centre at Shillong, who have also helped, advised and guided me from time to time.

I shall be failing in my duty if I do not acknowledge my gratefulness to all authors whose work I have consulted and quoted in this venture.

I would like to thank my friend Dr. L.P. Pathak, Librarian, and staff of the NEHU Library for their generous library support.

My thanks are due to all my friends in particular Professor N. Malla, Dr. R.P. Sharma, Dr. Veena Sharraf Chandel, Miss Therasia Dkhar and Mr. Joseph F. Khongbuh, who have helped me in giving final shape to this work.

Finally, I wish to acknowledge Mrs. Gayatri Pandey, my wife and my children, whose cheerful disposition supported me through the long months towards the completion of this work.

SHILLONG

MAY 10th 1996

(R. C. PANDEY)

Certificate i

Acknowledgements ii-iii

CONTENTS

		Pages
CHAPTER I	**INTRODUCTION**	1-7
1.1	Statement of the Problem	1
1.11	Objectives of the Study	1
1.12	Hypotheses	4
1.13	Methodology	5
1.131	Descriptive Approach	5
1.132	Analytical Approach	6
	Cited References	7
CHAPTER II	**LINGUISTIC ANALYSIS AND INFORMATION RETRIEVAL**	8-50
2.0	Introduction	8
2.1	Linguistic Analysis	8
2.11	Definition of Linguistic Analysis	8
2.12	Objectives of Linguistic Analysis	11
2.13	Levels of Linguistic Analysis	11
2.131	Phonology	13
2.132	Syntax	14
2.133	Semantics	17
2.14	Techniques of Linguistic Analysis	19
2.141	Immediate Constituent Analysis	20
2.142	Discourse Analysis	23
2.143	Componential Analysis	24
2.144	Distributional Analysis	28
2.145	Segmental Analysis	31
2.15	Generalization	32

Pages

2.2 Information Storage & Retrieval System 37
2.20 Introduction 37
2.21 Concept of Information Retrieval 38
2.22 Indexing Language 40
2.23 Structure of Relationships in Indexing Language 43
Cited References 48

CHAPTER III **PRINCIPLES OF INFORMATION RETRIEVAL BASED ON LINGUISTIC ANALYSIS** 51-172

3.0 Introduction 51
3.1 Principles propounded by different scholars 56
3.11 Charles Ammi Cutter 56
3.12 Julius Otto Kaiser 62
3.13 James Duff Brown 65
3.14 Henry Evelyn Bliss 69
3.15 Shiyali Ramamrita Ranganathan 72
3.16 Gerard Cordonnier 91
3.17 Jasua Edward Louis Farradane 93
3.18 Hans Peter Luhn 98
3.19 Mortimer Taube 102
3.20 Calvin N. Mooers 105
3.21 James W. Perry, Allen Kent and Thomas H. Rees 107
3.22 Douglas John Foskett 113
3.23 Barbara Kyle 123
3.24 Jean-Claude Gardin 126
3.25 Derek William Austin 129
3.26 Eugene Garfield 133
3.27 Eric J. Coates 136
3.28 B.C. Vickery 142
3.29 Jean Aitchison 145
3.30 Ganesh Bhattacharyya 149
3.2 Generalization 153
Cited References 166

		Pages
CHAPTER IV	**APPLICATION OF LINGUISTIC ANALYSIS IN INFORMATION RETRIEVAL TOOLS**	173-201
4.0	Introduction	173
4.1	PRECIS (Preserve Context Indexing System)	174
4.11	Syntax	174
4.12	Semantics	181
4.2	The UNESCO: IBE Education Thesaurus	186
4.21	Vocabulary Control	187
4.22	Thesaurus Structure and Relationships	189
4.221	Equivalence Relationships	189
4.222	Hierarchical Relationships	190
4.223	Associative Relationships	191
4.23	Thesaurus Display	192
4.24	IBE: Education Thesaurus	193
	Cited References	201
CHAPTER V	**CONCLUSIONS AND FINDINGS**	202-207
5.0	Conclusion & Findings	202
5.1	Conclusion	202
5.2	Suggestions	205
	Cited References	207
	BIBLIOGRAPHY	208-216

CHAPTER I

INTRODUCTION

INTRODUCTION

1.1. Statement of the Problem

There is no aspect of human experience and endeavour which is not touched, covered and affected by language. Language is not only a means of communication but at the same time, a storehouse of information. To the extent, Library and Information Science deals with classification, indexing, storing and dissemination of information, it comes very close to linguistics which studies the nature and function of language. In short, in a very significant sense, both linguistics and information retrieval are very close to each other because both are dealing with information. Further the development of artificial and formalized languages in the present century has strengthened our understanding of the deep structure of language and this has brought linguistics very close to computer technology and Robotics. As such all these developments in linguistics are bound to revolutionize library and information science.

The fundamental question of information retrieval is as: What is the basis of deciding how to represent

documents in a retrieval system? and even more fundamentally, is there a theoretical foundation which can support document representation procedure? The process of representing documents for retrieval is fundamentally a linguistic process, and the problem of describing documents for retrieval is, first and foremost, a problem of how language is used. Thus any theory of indexing or document representation presupposes a theory of language and meaning. Therefore, linguistic aspects of information retrieval are very vital. There is need to relate linguistics and information retrieval by considering the theories and techniques that linguistics has to offer, and how far these have been, or could be exploited by information scientists. Farradane also pointed out three main problems of analysis and organization of knowledge for retrieval. He says the processes of information storage and retrieval can be considered under three main heads: representing knowledge accurately whether it is the input to a system, or the question put to the system; organization of the recorded form of the information; and interpreting and satisfying the user's requirements. (Farradane 1970; 607). These problems are due to the nature of knowledge which is domain of psychology and linguistics.

Thus, it is clear that there is a close relationship, between the two fields of linguistics and information retrieval. Literature reviewed by (Garden 1973; Montegomery 1972; Spark Jones and Kay 1973) also prove that there is large common ground between linguistics and information retrieval, and this is further supported by the renewed interest in linguistics as a part of natural language processing applications in information retrieval.

Another factor which relates these two subjects together closely is, that the text in ordinary language is primarily concerned with information retrieval as well as linguistics. However, we find that there is lack of inter-penetration in these subjects of studies. There are various reasons that could be adduced for the lack of this communication between linguistics and information retrieval. One claim that might be made is that linguistics is still in a developing stage to provide a firm basis for the kinds of practicalities that information scientists are involved in. Now, question arises: Is there any possibility of changing this condition, and if so, what are the possibilities which can improve their relation?

1.11. Objectives of the Study

In the light of the above, the main objectives of the present study are as follows:

(i) To describe and assess the relationships and correspondence between linguistic analysis and information retrieval;

(ii) To trace out the ideas and techniques of linguistics having relevance for information retrieval;

(iii) To explore the possibility of a conceptual model for information retrieval based on linguistics; and

(iv) To evaluate the impact of linguistic ideas in information retrieval tools.

1.12. Hypotheses

The issues proposed to be examined in the context of the objectives of the study are as follows:

(i) Whether there is a correspondence between linguistic analysis and information retrieval at only two levels of linguistic analysis, i.e., Semantics and Syntax;

(ii) Whether principles of information retrieval propounded by different scholars at semantic and syntactic level bear upon the canons, principles and postulates developed by Ranganathan; and

(iii) Whether the conceptual framework developed by Ranganathan for syntax is loosing its grip in the field of information retrieval.

1.13. Methodology

To achieve the objectives of the present study, the following two methodological approaches have been adopted:

(i) Descriptive approach; and

(ii) Analytical approach.

The present study, being a conceptual study the methodological approaches like, historical, survey and experiment were found not to pay the dividends. Concepts are more amenable to descriptive and analytical approaches. This was the reason which led to the choice of these methods.

1.131. Descriptive Approach

The following have been covered under descriptive approach:

(a) Linguistics and Information Retrieval, their nature and correspondences;

(b) Study of relevant ideas useful for information retrieval; and

(c) Survey of principles as propounded by different scholars in the field of information retrieval based on linguistics.

1.132. Analytical Approach

The analytical approach has been used:

(a) To evaluate the principles of information retrieval based on linguistics;

(b) To search a conceptual model for information retrieval; and

(c) To evaluate the application of linguistic ideas in information retrieval tools.

CITED REFERENCES

Farradane, J.E.L. "Analysis and Organization of Knowledge for Retrieval", ASLIB PROCEEDING 22(2); 1970.

Gardin, Jean-Claude. "Document Analysis and Linguistic Theory", Journal of Documentation 29(2); 1973.

Montegomery, C.A. "Linguistic and Information Science", Journal of American Society and Information Science 23(3), 1972.

Spark Jones, K. and Kay, M. Linguistics and Information Science. New York: Academic, 1973.

CHAPTER II

LINGUISTIC ANALYSIS AND INFORMATION RETRIEVAL

LINGUISTIC ANALYSIS AND INFORMATION RETRIEVAL

2.0. Introduction

As is well known, the central task of Information Retrieval is to understand how documents should be represented for effective retrieval. It is claimed in the present study that the problem of representing documents for retrieval is basically a problem of describing linguistic facts of a given language. A theory of indexing or document representation should therefore presuppose a theory of linguistic analysis. The organization of linguistic facts and the procedures of linguistic analysis should reflect, at least in part, the inherent organization of data in a language and the means by which such data are processed by the human mind. Since it is assumed that the procedures of linguistic analysis and organization of linguistic facts may make information retrieval more readily accessible to users of informatics, it becomes important to first understand what "Linguistic Analysis" means.

2.1. LINGUISTIC ANALYSIS

2.11. Definition of Linguistic Analysis

Linguistic analysis may be defined as an organized attempt to describe linguistic data for a defined purpose.

As language behaviour is an integral part of human behaviour, it is but natural that structures, patterns and peculiarities occur again and again in any area of language use. A linguistic analysis thus seeks to identify and analyse such structures, patterns and peculiarities in order to make verifiable statements about a stretch of language use or a text. Linguists use "Linguistic Analysis" as a blanket term to refer to various types and techniques of such analysis. A linguistic operation performed by a linguist on the data of a text is linguistic analysis. Such data may be collected from field work or from a recorded material, whether spoken or written. There are operations, types or techniques of Linguistic Analysis but it is significant to note that what all these operations, types and techniques have in common is the implicit belief in the regularity or recurrence of certain structures, patterns and peculiarities.

When a linguist begins to perform his operations on received linguistic data, he has to break:

> it down into minimal elements (segmental analysis), by establishing how these elements are made up (componential analysis), by examining the relationships between them (distributional analysis) and by determining their arrangement within longer sequences (immediate constituent analysis, functional analysis). Such procedures are used on the levels

> of sound (phonemic analysis), grammar (grammatical analysis, morphological analysis) and vocabulary (semantic analysis). (Hartman and Stork 1972: 130)

Although no analysis or description of a language can hope to explain or describe everything that the language consists of or refer to, one has to formulate a system of analysing and describing a language one is interested in. A language is not only a system but is a human system, therefore it is the "human" part of the system that makes it a unique phenomenon. Any linguist who undertakes linguistic analysis is or should be aware of the difficulties of the task he is engaging himself in. This should not, however, mean that language are unanalysable or indescribable. Languages can and must be analysed and described. They are, in fact, analyzed and described in spite of the awareness of such limitations. Three components of language are universally recognized. They are: phonological, syntactic and semantic. (Katz and Postal 1964: 1; Chomsky 1967: 16). In other words, language has sounds, structures and meanings. A linguistic analysis is an attempt to testify these components with reference to a particular language or a group of languages.

2.12. Objectives of Linguistic Analysis

If we care to analyze today how a language works, we may be tempted to suppose that it must have originated at a certain stage of evolution of a particular community. It is almost needless to say that a language works for communication between the members of the community which uses that language. The purpose of describing or analysing a language, however, is not communication in this sense. The purpose of Linguistic Analysis is to discover identifiable facts about a language. Linguists analyze a language to learn about a language as well as to share their knowledge of that language with like-minded students and researchers of language. The general objective of linguistic analysis is, therefore, to develop procedures which make it possible to process information content about a language.

2.13. Levels of Linguistic Analysis

One can easily identify several levels of linguistic analysis from graphetics to semantics. Some of these identifiable levels are:

(i) Graphetics and graphology;

(ii) Phonetics and phonology;

(iii) Morphology;

(iv) Syntax and grammar; and

(v) Semantics.

However, if we take a close look at these levels, we should be able to select three crucial levels. These levels could be: Phonology, syntax and semantics. To say that a language consists of these three components or to say that a linguistic analysis must be confined to these three levels would certainly be an over-simplication. But as the present study does not attempt an indepth study in the area of linguistics, it can take a general view of the levels that linguistic analysis is ordinarily concerned with. The present study is based on the framework used by R.W. Langacker in his discussion of these levels in the first part of the opening chapter of his well known book, Fundamentals of Linguistic Analysis.

Langacker maintains that a language, "is a series of correlations between the sounds and meanings achieved through a networking of rules and structures" (1972:3). A language is systematically structured and rule-governed. A language is a set of principles or rules that control how speech sounds become words, how words become sentences, and how sentences acquire meaning. These processes are called 'structures' and structures are by definition regular and predictable. Rules or principles and structures join hands to make a language operational with

the components of phonemes, words, sentences and meaning. "A rule or principle", says Langacker, "is a regularity in the structure of a language" (1972:3). One can begin with phonology and end up with semantics. Although Langacker begins with meanings and finally comes to the analysis of sounds. For our convenience, it can be done in the opposite way.

2.131. Phonology

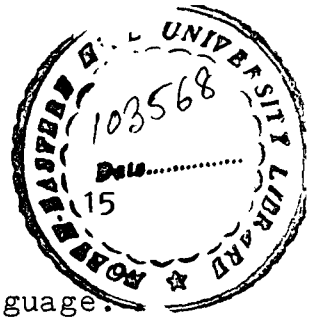
As we all know, Phonology is concerned with the patterns and organization of a language in terms of its phonetic features and categories involved. The smallest unit of phonology is called 'Phoneme'. A language has a finite number of phonemes. English, for example, has 44 phonemes (20 vowels and 24 consonants). No word in English can consist of a phoneme or of a group of phonemes which do not form part of the 44 English phonemes. There are rules for combining phonemes to make words. These rules describe the patterns and organization of phonemes in the English language. Any linguistic analysis that attempts to describe the English language has therefore to identify, describe and analyze these 44 phonemes and rules which enable these phonemes to form words.

Phonemes and allophones are generally distinguished. A Phoneme is a class of phonetically similar sounds,

contrasting and mutually exclusive with all similar classes of sounds in the language. The individual sounds which compose a distinctive phoneme sound are its allophones. Every occurrence of 'P' for example, is an allophone of the phoneme 'P'. A phoneme is a family name for a certain class of sounds. Various sounds constitute a phoneme like 'P' which is minimal, distinctive and exclusive, compared to another phoneme like 't' or 'g'. This means that all the sounds which are classifiable as 'P' are the allophones of the phoneme 'P'. A phonemic analysis of a language is competent to identify the phonemes of the language. Since phonemes are the smallest units of a language, no language analysis can ignore to take note of these infrastructural elements of the language in question.

2.132. Syntax

Traditional syntactic analysis works with two fundamental units of description: the word and the sentence. Now the question arises, 'What is a word?' Is 'black-bird' one word or two. If 'accept' is a word, is 'unacceptable' also a single word? There is no satisfactory answer to such questions. In order to solve such problems linguists have identified a unit smaller than the word: the 'morpheme'. A morpheme is the smallest



meaningful unit in the grammatical system of a language. As language consists of phonemes, it consists of morphemes. Morphemes are of two types: free morphemes and bound morphemes. Free morphemes are words such as, 'cat', 'pen', 'man', 'book' etc. A word, however, is not necessarily a free morpheme. Some words are free morphemes. But other words are clusters of free and bound morphemes. A word is either a free morpheme, a combination of free morphemes or of a free and bound morphemes. 'Man' is a word because 'man' is a free morpheme 'man-like' is a word which is a combination of two free morphemes (man and - like), 'manly' is a word which combines the free morpheme 'man' with a bound morpheme '-ly' while 'manliness' is a word which combines the free morpheme 'man' with the bound morphemes '-ly' and '-ness'.

Looking at the word from the structural point of view, one is generally inclined to say that a free morpheme can stand on its own, while a bound morpheme cannot. But the question is how free is the 'free' morpheme or the word constituted of it. If we pause for a moment and try to analyze the observation just made, we will soon discover that full words do not occur freely. A language primarily consists of larger constructions which may be called 'sentences'. It is immaterial whether

a sentence is short or long. It is possible to have a sentence, which consists of a single word or free morpheme, which in turn consists of a single phoneme. The single phoneme 'I' is a free morpheme, which is a word and which can easily become a sentence. Grammarians tell us that sentences need not have both subjects and objects at a time. A sentence can consist of a single phoneme, which is a single morpheme and a single word on the one hand, while on the other a sentence can consist any number of words. There are simply no rules about the maximum number of words in a sentence. When we speak of a language then, we speak of the infinite length and variety of sentences of the language. When we listen to another language, the first thing that strikes us is a jumble of sentences, phrases, words; morphemes and phonemes come later on. No complete linguistic analysis can ignore the sentences of a language. Syntax is the name which refers to the actual constructions of a language. The syntactic level of linguistic analysis, therefore, acquires a unique significance in the study of a language. Noam Chomsky points out that 'competence' and 'performance' of the native speaker are important issues to be taken in syntactic analysis. The native speaker of a language seems to possess an internalized set of rules of sentence construction. It is by virtue of his knowledge

of this set of rules that he can understand and produce constructions as sentences in his language. The native speaker's ability to understand sentences relates to his 'competence' while his ability to produce sentences relates to his 'performance' (Crystal 1980: 315-6).

Chomsky also distinguishes between deep structures and surface structures in a sentence. Each sentence is considered to have both: the 'deep structure' and the 'surface structure'. "The surface structure is generally the syntactic structure of a sentence which a person speaks, hears, reads or writes", while the deep structure "is much more abstract and is considered to be in the speeaker's, writer's, hearer's or reader's mind" (Richards 1985: 74). 'John is easy to please' and 'John is eager to please' are superficially similar in structure but in fact, very different in meaning. The surface similarity between sentences can be deceptive with regard to their meaning. The speaker of these structures, Chomsky would say, wants to convey very different meanings even though he uses apparently similar syntactic structures. The analysis of the language at the syntactic level must pay attention to such matters of philosophical importance.

2.133. Semantics

Words in a language, as has been said, are impor-

tant, but sentences in a larger context are more important. Sentences are important for their form. But they acquire greater importance when they are recognized as tools to convey meanings. Sentences have a beginning but they have more importantly an end. Meaning becomes the ultimate end in a sentence. Sentences are valuable only because they represent some meaning. It is impossible to think of important sentences which do not mean. They are important only when they mean and succeed in conveying their meaning with least resistance. The study of sentences from the point of view of meaning takes us to the semantic level of linguistic analysis.

Semantics is devoted to the study of meaning in language. In linguistics the study of meaning is generally confined to the identification and analysis of the semantic properties of natural languages. Terms like 'behaviourist semantics' and 'structural semantics' are sometimes used to refer to the study of meaning through the notion of semantic relations in 'generative linguistics'. This term refers to the study of the meaning of sentences. Semantics has various fields and techniques, which are often exclusive of each other. From Bloomfield to the present day, it has undergone several transformations but most linguists now seem to agree that the seman-

tic component of a language should form part of any worthwhile linguistic analysis of that language (Crystal 1980: 315-6; Richards 1984: 254).

2.14. Techniques of Linguistic Analysis

There are different ways of analyzing the facts of a particular language. Various schools of linguistics attempto to define, describe and explain language from their respective points of view. The approach of a linguist (or a school of linguistics) determines the kind of things the linguist has to say by way of analyzing his corpus. It often happens that two linguists dealing with the same corpus have to say two different things to an extent that it looks like talking about two different corpora. This might happen so mainly because the two linguists have two entirely different approached.

Linguistic analysis may be conveniently divided into different kinds or types with reference to various schools of linguistics. One can speak of linguistic analysis in terms of structural grammar, T.G. grammar, Tagmemic grammar, Systemic grammar, Case grammar, Stratificational grammar and the like. It is not important for the present study to find out whether a particular approach is more or less sophisticated than the other from the linguistic point of view. What is important for this study is whether

a particular way, procedure or technique of linguistic analysis is more useful than another way, procedure or approach from the point of view of information retrieval. It is for this reason that references to some well known techniques of linguistic analysis are classified and briefly discussed. This classification is based on Hartmann and Stork's classification of the techniques of linguistic analysis. They classify the techniques of linguistic analysis as follows:

> Linguistic analysis is a collective term for various operations that the linguist performs on data he has obtained from field work or by collecting texts. Linguists attempt to find the regular patterns in this material by breaking it down into minimal elements (segmental analysis), by establishing how these elements are made up (componential analysis), by examining the relationships between them (distributional analysis, discourse analysis), and by determining their arrangement with longer sequences (immediate constituent analysis, functional analysis). Such procedures are used on the levels of sound (phonemic analysis), grammar (morphological analysis, grammatical analysis), and vocabulary (semantic analysis). (Hartmann and Stork 1972: 130).

2.141. Immediate Constituent Analysis

The linguist is interested not only in sentence patterns, which really represent the sequence of sentence elements (determiner, noun, verb etc.) but also in the

relationships between the elements. Which elements are more closely related to each other? In the sentence, 'The hunter shot the tiger', which of the following ways of grouping the elements (words) is acceptable to a native speaker?

(i) The / hunter shot the tiger.

(ii) The hunter / shot the tiger.

(iii) The hunter shot / the tiger.

(iv) The hunter shot the / tiger.

A native speaker would intuitively say that he (ii) identifies the two main parts of the sentence, i.e., the parts that make up the sentence "immediately" (Parts that immediately make up a sentence or larger parts of a sentence are called "immediate constituents". The immediate constituents of the above sentence are 'the hunter' and 'shot the tiger'. The linguist can continue the process of dividing the sentence into parts in order to find the groupings acceptable to a native speaker. For example, 'shot the tiger' is further divided into immediate constituents like 'shot' and 'the tiger'. The process can be continued until only individual words are left.

The above process is called "immediate constituent

analysis" or IC analysis. IC analysis may be diagrammed in various ways, two of which are given below:

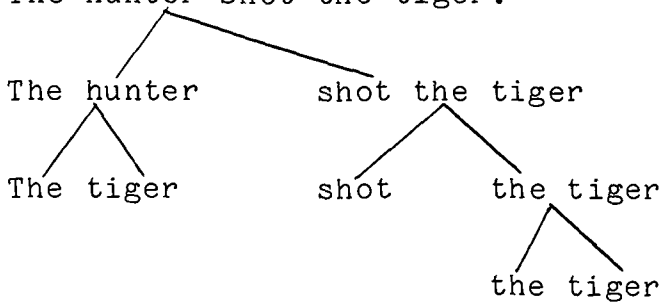
A. The hunter shot the tiger.

(i) The hunter/|shot the tiger

(ii) The/|hunter/|shot/ the | tiger

(iii) / the/| tiger

B. The hunter shot the tiger.



The diagrams make two things clear: one is that IC analysis involves a process of binary division (dividing into two parts each time). The other is that language is not made up of a string of elements in sequence, for the divisions are ordered in importance within a simple hierarchy. Traditional grammarians practised a form of syntactic analysis called 'parsing' i.e., the identification of parts of a sentence as subject, verb, object etc. and of words in a sentence as noun (plural) verb (past tense etc.). Bloomfield kept the same type of analysis, but substituted formal criteria for the logical ones, thus creating immediate constituent analysis. The

procedure was originally set forth by Bloomfield, provided with exegesis by Pike and systematized by Well (Paillet and Dugas 1982: 101-2).

2.142. Discourse Analysis

In the study of language, two things are important; the way a language is used, and the way its components are combined. When we try to investigate how, we, as language-users make sense of what we read in texts, understand what speakers mean by what they say, recognize as connected (as opposed to) jumbled or incoherent discourse, and successfully take part in that complex activity called conversation, we are undertaking what is known as 'discourse analysis'.

Discourse is a general term, for example, of language use, i.e., language which has been produced as a result of an act of communication. According to Panini, three conditions are essential for a discourse, "it must contain knowledge about something outside itself, it must be well-ordered, and it must receive social legitimacy as a system of knowledge" (Quotes in Devy 1992: 89).

The study of both written and spoken discourse is known as 'discourse analysis'. Jack Richards (1985:84)

defines Discourse Analysis as "the study of how sentences in spoken and written language form larger meaningful units such as paragraphs, conversations, interviews etc." For example, discourse analysis deals with:

(a) How the choice of articles, pronouns and tenses affect the structure of the discourse;
(b) The relationship between utterances in a discourse; and
(c) The moves made up by speakers to introduce a new topic, change the topic or assert a high role relationship to the other participants.

Thus, we can say that 'Discourse Analysis' studies the use of language in discourse, the manner in which the discourse content of any subject is organized and given linguistic expression. The aim is to reveal the process by which a piece of language is interpreted as discourse. We try to discover linguistic regularities in a discourse, using, phonological, grammatical and semantic criteria (e.g., cohesion, anaphora, inter-sentence connective).

2.143. Componential Analysis

What is meant by the term componential analysis

in semantics is being explained by means of the simple example that has been often used for this purpose by linguists:

(i) man - woman - child

(ii) bull - cow - calf

(iii) stallion - mare - foal

On the basis of our intuitive appreciation of the sense of these words, we can set up such proportional equations as the following:

Man : Woman : Child :: Bull : Cow : Calf

This equation expresses the fact that, from the semantic point of view, the words 'man', 'woman' and 'child', on the one hand, and 'bull', 'cow' and 'calf', on the other, all have something in common. Furthermore, 'bull' and 'man' have something in common which is not shared by either cow or woman or calf and child; that cow and woman have something in common which is not shared by either bull and man or calf and child; that calf and child have something in common that is not shared by either bull and man or cow and woman. What these different groups of words have in common is called a semantic component, (other terms have also been used in the literature, plereme, sememe, semantic maker and semantic category etc.)

The componential approach to semantics is inherent in the traditional method of definition by dividing a genus into species and species into sub-species and this method of definition is reflected in most of the dictionaries that have ever been compiled for particular languages.

In componential approach to semantics, one can answer two different questions with the same set of components. The first question has to do with the semantic acceptability of syntagmatic combinations of words and phrases: whether a given combination is to be generated as significant or excluded as meaningless? The second question is: What is the meaning of a particular combination of lexical items?

The significance of grammatically well formed sentences is traditionally accounted for in terms of certain general principles of 'compatibility' between the meaning of their constituent lexical items. One way of expressing this notion of semantic 'compatibility' is to say that the relevant semantic components of the lexical items in the syntagmatic combination generated by the syntax must not be contradictory. For example, the word 'pregnant' contains a component which restricts it to the modification of nouns which contain the

component (female); on the basis of the fact (modification being interpreted by the syntactic rules of the language) such phrases as the 'pregnant woman' or 'a pregnant mare' would be generated as significant, and such phrases as 'a pregnant man' or 'a pregnant stallion' would be excluded as meaningless.

The second question that Componential Analysis sets out to answer is, what meaning does a given sentence or phrase have? The answer to this question is that the meaning of a sentence or a phrase is the 'product' of the senses of its constituent lexical items, and the sense of each lexical item is the product of its constituent semantic components. The meaning of a sentence or a phrase is therefore determined by an 'amalgamation' of all the semantic components of the lexical items according to a set of 'projection rules'. The projective device consists of two parts as propounded by Katz and Fodor: first, a dictionary that provides a meaning for each of the lexical items of the language, and second, a finite set of projection rules. The projection rules of the semantic component assign a semantic interpretation of each string of formative generated by the syntactic component. The semantic interpretation that a string of formatives has assigned to it provides a full analysis of

its cognitive meaning. Thus we can say that componential analysis is a technique for the economic statement of certain semantic relations between lexical items and between sentences containing them.

2.144. Distributional Analysis

A language basically consists of two parts: form and meaning. Therefore, it is essential to make a clear-cut distinction between formal analysis and analysis based on meaning while making analysis of a language.

In the formal approach of a linguistic analysis, the main aim is to establish a set of units, which are describable, "in physical terms of form, correlation of these forms and arrangement of order" (Fries 1952: 58).

In the spoken or written language the individual sounds, words, and longer combinations occur only in a restricted number of combinations. It is possible, therefore, to define linguistic units in terms of their *distribution or "privileges of occurrence" in relation* to each other. For doing this, we can specify a particular environment in which items can occur. A linguistic environment is a selected point in the linear representation

of language. According to Hartmann & Stork,

> Distribution is the number of possible environments or contexts in which a particular linguistic item such as speech, sound or word can occur in a given language, dialect or portion of text.

He further defines distributional analysis as:

> A method of linguistic analysis which shows the distribution of phonological, grammatical or lexical elements within larger sequence, e.g., phones in words or words in sentences. The frequency of occurrence of individual units in restricted environments is considered here as important as their functional interrelationships - functional analysers. (Hartmann & Stork 1972: 71)

This type of analysis is based on the fact that every linguistic unit below the level of the sentence has a characteristic distribution; that is, it is restricted to a greater or lesser degree with respect to the environment in which it occurs. Two or more units occurring in the same range of contexts are said to be distributionally equivalent; if they never occur in the same context they are complementary. There are certain intermediate cases where the distribution of one unit includes the distribution of another without being totally equivalent to it, or where the distribution of two units overlaps but without either of the two occurring in all

the contexts where the other occurs. For example, in the English word "Pan', we can consider the '-an' part of the word to be the linguistic environment in which the sound 'p' occurs. All of the consonants of English can substitute for 'p' to form a normal word, but not the initial sound in 'thigh', 'high', 'why', 'yes', 'zoo' and the final sounds in 'rouge' and 'sing' are in complementary distribution with all the other sounds that occur there. In other words, we can say that all the other sounds that can be substituted in this environment have a differential function in that position, while the others do not.

A formal analysis of language reveals the typical constructions, the language uses, the member that can and cannot form part of these constructions and the characteristics of construction that contrast with each other. Such an analysis assists us in determining the function of units, patterns and contrasts of the language whether differential, referential or both.

The principles of distributional analysis were most thoroughly and extensively discussed by Harris (1951) especially in, "Methods in Structural Linguitics". His intention in describing this analysis was to formulate

a technique or procedure, which could be applied to a corpus of utterance and, with the minimum use of informant's judgements of 'sameness' and 'difference' could be guaranteed to derive the rules of the grammar from the corpus itself. Latr on, these principles of distributional analysis and its techniques propounded by Harris and his colleagues served as the foundation upon which Chomsky developed his generative grammar.

2.145. Segmental Analysis

When an utterance is analysed into parts so that each part corresponds to a given stretch of sound (or sequence of letters), these parts are called segments of the uttrance. 'Segment' is a term in linguistics which refers:

> To any discrete unit that can be identified, either physically or auditorily in the stream of speech. Segmentation can take place using either physical or auditory criteria in the former case, acoustic or articulatory change points can be identified as boundaries of segments, in the latter case, perceptible change in quality or quantity, often showing the influence of the language phonemic units are the basic of division. (Crystal 1980: 314).

In phonology, a major division can be made into two parts: Segmental and supra-segmental. Segmental

phonology analyses the speech into distinctive units, which have direct correspondence with phonetic segments.

Supra-segmental phonology analyses those features of speech which extend over more than one segment, such as intonation, or the Vowel-Consonant structure, of the formatives in surface structure. Segmental analysis is defined as, "a method of linguistic analysis into units e.g., Phonemes or Morphemes, which in turn may be used to construct larger units, e.g., words or sentences" (Hartmann & Stork 1972: 202). For example, an English word 'dehumanization' has the following morphemes:

(i) de-
(ii) -human
(iii) -ize
(iv) -ation.

Hjelmslev explicitly conceives of the linguistic object as a set of texts in which, by segmentation, the regularities of grammar have to be found. The regularities thus discovered permit deductively, a predictive description of all the possible texts of a language.

2.15. Generalization

It may have become clear from the foregoing discussion that linguistic analysis operates on various

levels and uses various techniques. These levels and techniques cover the whole gamut of linguistic activity from the individual speech sounds to the largest discourse and their meanings. The place and manner of articulation of speech sounds by the speaker and their reception and comprehension by the listener are very important activities indeed. The way syllables, words and sentences are written by the user and the way they are read and understood by the reader are also very important linguistic activities. Just as levels like graphetics and graphology, phonetics and phonology are universally recognized as basic insights into the nature, behaviour and function of a human language. Similarly, segmentation and distribution constitute the basic varieties of linguistic analysis. But if we examine these levels and techniques from the point of view of the information retrieval system, we soon realise that despite their great importance in linguistic scholarship, these levels and techniques do not help the information scientists much for the simple reason that information science is particularly concerned with the acquisition, organization and dissemination of information through language which begins with words and sentences and ends with their meaning(s). Even if it tries to be very specialized in its use of language,

an information system cannot make much practical use of language at the phonetic/phonological level nor can it avail of segmentation and distribution of speech sounds in its day-to-day use of language. Likewise, it cannot make much use of the knowledge of how a piece of language is written by a particular writer or received by a particular reader.

Information science is practically concerned with language which consists of discourses, with sentences which consist of phrases and words as well as with meanings of words, phrases, sentences, discourses and works. It has, therefore, to be concerned with those levels, categories, tools and techniques, which describe language from words to the discourse or text. Techniques of syntactic and semantic analysis therefore acquire great significance for information retrieval system.

If we look at the different techniques of linguistic analysis with reference to the pragmatics of information system, we come to the conclusion that techniques like IC analysis, componential analysis and discourse analysis constitute the three important areas of linguistic scholarship that a student of information science can greatly benefit from. Scholarship in linguistics has

been evolving from age to age. Linguists have been formulating and replacing the techniques of linguistic analysis from time to time. No technique of linguistic analysis has yet been evolved which could answer all questions or meet all the requirements of an adequate or foolproof analysis of a language. Each technique has its own share of limitations. It is, therefore, important to note here that when we refer to some techniques of linguistic analysis as useful techniques, we do not lose sight of their inadequacies. What we actually intend to do is to point out that such techniques are useful tools for the students of information system, in spite of their inadequacies within the science of language proper. Information system is in no position to pass judgements on the merit of a particular development in linguistics. It can only make use of what is available in linguistics for its own purposes. It is from the point of view of such purposes that we find that IC analysis, Componential Analysis, Discourse Analysis, Segmental and Distributional Analysis are useful tools for identifying, classifying and interpreting linguistic data. IC analysis is basically a technique of syntactic analysis. The aim of IC analysis is to analyse each utterance and each constituent into maximally independent sequences by dividing constituents

of a given binary unit. It describes what constituents of a given utterance are and helps us to discover how constituents are layered in sentences, each lower-level constituent being part of a higher-level constituents.

If we further analyze these above mentioned techniques, we find that these are parts of basically two levels of linguistic analysis: syntactic and semantic. The process of content analysis in information retrieval involves the identification of the concepts contained in information records and requires the determination of the relations linking these concepts. The first process needs (was) some kind of semantic analysis and the second syntactic analysis. These two types of analysis are highly essential to information retrieval. From linguistic analysis point of view, IC analysis, Distributional Analysis, and Segmental analysis are parts of syntactic analysis, whereas Componential analysis and discourse analysis are parts of semantic analysis. Therefore, now onward, we will concentrate our attention on these two levels of linguistic analysis: syntax and semantics and try to trace out contributions of different information scientists in the light of these two levels in the rest of the thesis.

2.2. Information Storage & Retrieval System

2.20. Introduction

So far we have been dealing with language, its levels and techniques which are useful from the point of view of information retrieval. It is time to look into how language is helpful to make information retrieval effective and efficient. Before going into the detailed study of the relationship, between language and information retrieval, it will be necessary to understand the idea underlying the term 'information retrieval'.

Library science is, in essence, the retrieval and dissemination of embodied knowledge to individual members and groups in a community. Ranganathan has also emphasized the same idea in his first law of library science i.e., 'Books/Information are for use', here, he explains the key words of his first law 'book/information' as a generic name to denote all kinds of documents and the term 'use' as the use of organized, expressed and embodied knowledge, that is, the subject dealt with in documents - by the users. A document or the information retrieval system is, therefore, essentially concerned with the classification search, retrieval and service of subject. (Neelameghan 1967: 337-8).

2.21. Concept of Information Retrieval

The term 'Information Retrieval' was coined by Calvin Mooers in 1950 to solve the problem of "searching and retrieval of information from storage according to specification by subject" (Vickery 1968: VII). Somehow and for some reason, such as being assigned to a new project or work, the user senses a lack of information. Belkin (1980: 133-43) calls this an "anomalous state of knowledge" (or ASK), a gap or discontinuity in the person's knowledge. In other words, the person recognizes the lack of or need for data, information or knowledge.

The basic aim of information retrieval is to put readers into contact with right text information to rectify the gaps and anomalies in their personal states of knowledge. In other words, information storage and retrieval system must provide users with documents which start from their particular level of understanding and contribute to their specific problem - solving needs.

Information Storage and Retrieval System (ISPR), as it appears by name, has two aspects:

> Storage and organization of information, on the one hand, and the retrieval, on the other. These two aspects cannot be separated from each other: retrieval requirements dictate how the store should

> be organized; conversely, the organization of the store determines its usability for retrieval, its search ability. (Soergel 1985: 7)

According to Karen Spark Jones, an information retrieval system consists of:

> A file of documents with brief content characterizations in the form of index descriptions in an indexing language; a procedure for searching the file given the index description of a user's request for documents derived from his need for information about something; and a matching criterion for evaluating comparisons between request and document descriptions. The system also includes procedures for deriving the index descriptions of documents, and requests. It is driven by the user's need and the consequent assessment of retrieval documents for their relevance to this need. Cf. (Jones 1990:1)

A careful observation of the definition given by Jones reveals that information retrieval is a complex phenomenon which consists of documents, request, shorthand description of these documents and requests, a mechanism to match these descriptions and people. According to F.W. Lancaster informationn retrieval:

> Encompasses all activities involved in the storage and retrieval from the time a document is indexed for input to the system until it is retrieved and delivered to a user in response to a request made to the system. Cf. (Lancaster 1968: 5)

Soergel tried to make information storage and retrieval system more clear while describing ISAR system as a sub-system of total information system. He analyses each part of ISAR with the help of a diagram. According to the structure of this diagram, ISAR:

> Has two inputs; (i) query statements (descriptions of the needed information or entities couched in natural language) and (ii) entities or information about entities as acquired. The output of the ISAR system consist of entities or information about entities and their relationships presumed relevant for a query that is helpful for solving the problem that gave rise to the query. (Cf. (Soergel 1985: 57).

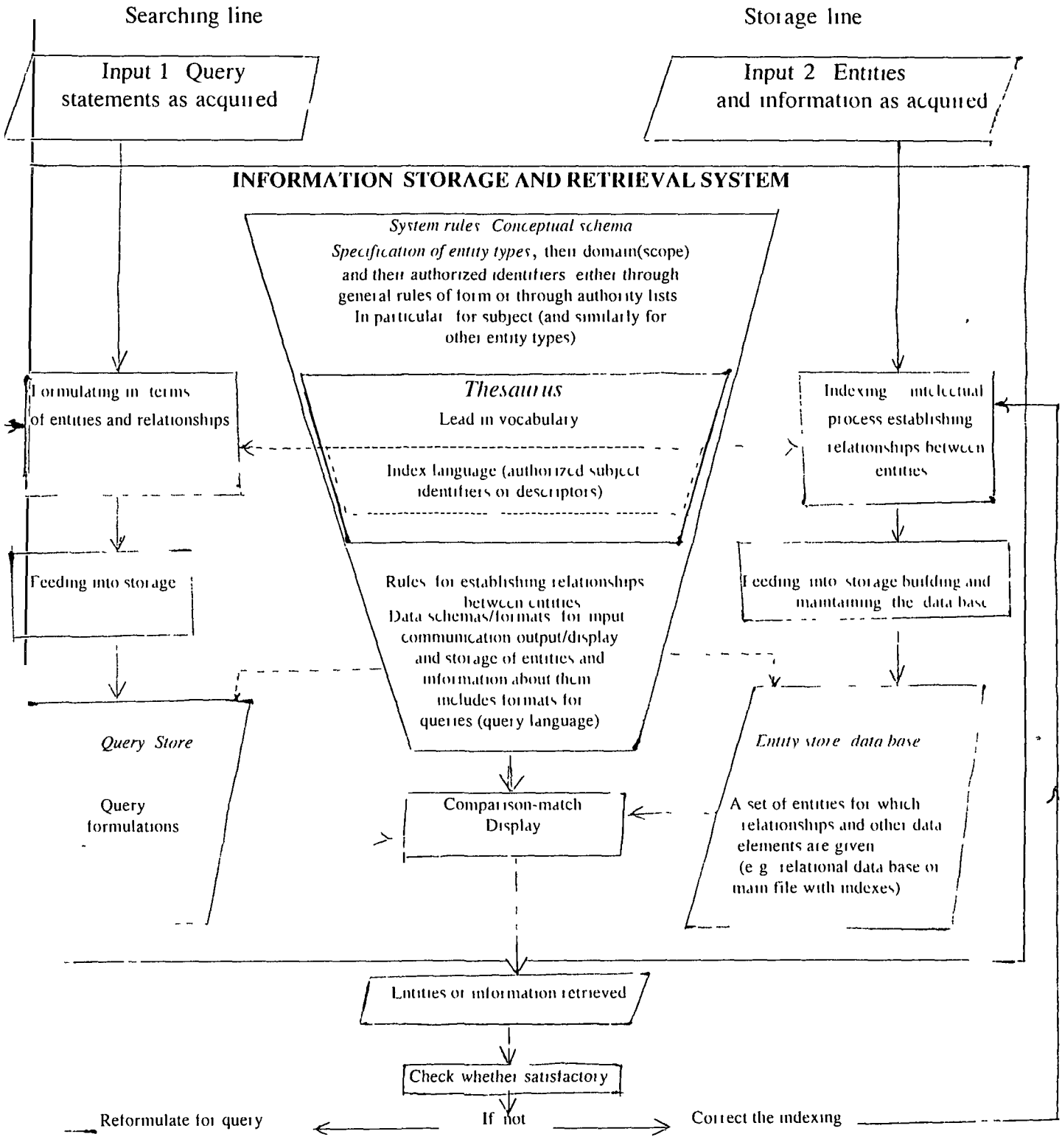
He further goes on elaborating his structure of the ISAR and tries to establish that:

> The entities and relationships included in the data base can now be matched against the query formulation and the entities or information meeting the specification extracted; the result is a data base - a mini data base, as it were displayed and formatted according to the specifications of the query formulation. (Soergel 1985: 59)

2.22. Indexing Language

Effectiveness and efficiency of any retrieval function is dependent on its specificity, exhaustivity, speed, ease and degree of satisfaction of the user. To

The Structure of Information Systems



The Structure of an ISAR System

achieve these aims, it is essential to adopt a mechanism which consists of highly generalized, hospitable, versatile and, therefore, adaptable intellectual foundation for all manifestation of information retrieval. The efficient and working mechanism in the field of information retrieval is the artificial language, which is also known as "Subject indexing language." In short, we call it "indexing language."

It is customary to refer to the basic units of document and request description as 'terms'. These are normally word like, and if extracted may be called 'key words' and while compound may be allowed, it is more usual to refer to 'complex multi-word descriptive units as subject heading'. The indexing vocabulary may then consist of the basic terms, which if controlled may be called 'descriptors', or 'subject heading'. The vocabulary itself may be given an organized retrieval structure as a thesaurus, list of subject headings or classification. According to Lancaster:

> Any query of user's is specific and therefore, to provide the information against such queries subject index or subject indexing is essential. He illustrates this in this way, when we use the term information retrieval system, however, we think primarily of a system capable of retrieving documents in answer to

> particular subject requests. The principal index is, therefore, the 'subject index' and the technique of entering documents into such an index is 'subject indexing'. (Lancaster 1968: 3)

Further, he says that:

> When a controlled vocabulary is set up in the form of an alphabetical listing of index terms, the individual term are known as subject headings and the controlled vocabulary as a list of subject headings. When on the other hand, the class labels are organized systematically so that related terms are brought together, and specific terms are shown subordinated to their appropriate generic terms, the controlled vocabulary is known as classification schedule. (Lancaster 1968: 7)

Again, he concludes:

> That it is important, however, to recognize that the activity of subject indexing remains the same whatever the form of controlled vocabulary used. In all cases we classify or group documents by assigning index term to them. (Lancaster 1968: 8)

While defining indexing language, Soergel expresses the same view. According to him, "index language is commonly understood to imply a controlled vocabulary. The same is true for terms system vocabulary and classification (Classification scheme), which are quasi-synonymous of index language" (Soergel 1985: 221). Neelameghan

explaining the importance of vocabulary control device expresses the same point of view, where he says that:

> The need of vocabulary control in indexing system - manual and machine based - is well recognized. The control may be at the input stage for the representation and organization of subjects in the data base or the retrieval stage in the expression of user's query or at both stages. Scheme for subject classification and thesaurus are among the commonly used tools for vocabulary control in indexing systems. (Neelameghan and Maitra 1977: 1-2)

It is clear from the above discussion that the foundation of information retrieval is subject analysis and subject analysis is the basis of construction of vocabulary control device like classification scheme or thesaurus, with a set of rules of syntax for its structuring of subject statement and formulation of query.

2.23. Structure of Relationships in Indexing Language

A structured subject statement consists of two things: terms and pattern of relationship. We know that terms stand for concepts and concepts combine in the statement according to a pattern of relationship. According to French linguist Ferdinand de Saussures, "The signs (term) that make up language are not abstractions, but real object, signs and their relationships are what

linguistic studies" (D.J. Foskett 1970: 25). Soergel also shares the same idea and mentions that, "there are many relationships among concepts, that make up the index language structure. This structure serves essential functions in indexing, data base organization, and searching" (Soergel 1985: 251). He categorises these relationships in the following way:

(a) Hierarchy, and

(b) Concept combination.

W. J. Hutchins described three types of relationships (Hutchins 1975: Sec 7.1):

(a) Paradigmatic relations such as 'genus species' relation, 'whole-part' relations, antonymy etc.

(b) Syntagmatic relations (Associative relations).

(c) Free syntagmatic relations.

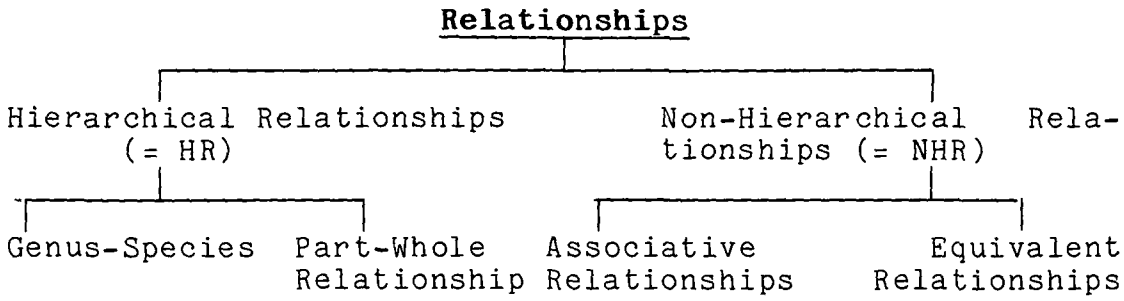
Bhattacharya categorises structural relationship of a subject proposition in three parts (Bhattacharya 1981: BA 23):

(i) Semantic structure, 'Genus-speciees', 'whole-part', and inter-facet relationships.

(ii) Elementary structure - It means the dimension of elementary categories to which the constituent concepts of a subject - statement belong.

(iii) Syntactic structure - Sequence of constituent concepts in a subject statement.

Neelameghan has explained structure of subject statement in the following chart:



If we look up hierarchical relationships in thesaurus, we find its further expansion as narrower term = (NR)/Broader terms (= BT) and Non-hierarchical relationships are represented as related terms = (RT).

Neelameghan is of the opinion that, "every type of NHR can be represented by one or the other of the following relations – facet relation, speciator relation, phase relation and coordinate relation in the facet analysed representation of the subject" (Neelameghan

& Maitra 1977: 3). Derek Austin explains two types of relation in the structure of subject statement as, "One operating at the level of concept, and the other concerned with the combination of concepts. He labels first of these as semantic or thesaural relationships and the second as syntactic relationships" (Austin Derek 1984: 199).

A.C. Foskett has also expressed two types of relationship in the structure of subject statements. According to him:

> There are two kinds of relationships between terms that we have to take account of, the recognition of terms denoting related subjects such as water, sea, & river, and the association or coordination of otherwise unrelated terms to represent composite subjects. The first kind are known as semantic relationships, while the second are called syntactic. (Foskett, A.C. 1982: 72).

Thus, we can conclude our discussion that an information retrieval system needs to have structured indexing language, which consists of terms available in the structure of subject statements and relationship between them. In other words, we can say that indexing language is highly essential to obtain the objectives of IR. An indexing language consists of concepts and

relationships among concepts. Relationship identified so far have been categorised into two parts: semantic and syntactic. Now, we will try to search for conceptual framework of semantic and syntactic relations in the principles as propounded by different scholars in the field of Library and Information Science.

CITED REFERENCES

Austin, D. A Manual of Concept Analysis and Subject Indexing, 2nd Edition, London: British Library, 1984.

Bhattacharyya, G. "Subject Indexing Language: Its Theory and Practice," DRTC Refresher Seminar, V.13, 1981.

Bhattacharyya, G. "Role of Classification and Indexing in Information Retrieval," DRTC Annual Seminar, V. 20, 1983.

Belkin, N.J. "Anomalous States of Knowledge as a Basis for Information Retrieval". Canadian Journal of Information Science, V. 6, 1980.

Chomsky, Noam. Syntactic Structure. The Hague: Mouton Pub., 1957.

Crystal, David. A First Dictionary of Linguistics and Phonetics. London: Andre Deutsch, 1980.

Davy, G.N. After Amnesia; Traditional and Change in Indian Literature Criticism. Hyderabad: Orient Longman, 1992.

Foskett, A.C. The Subject Approach to Information. 4th Edition, London: Clive Bingley, 1982.

Foskett, D.J. Classification for a General Index Language, London: The Library Association, 1970.

Fries, C.C. The Structure of English: An Introduction to the Construction of English Sentences, New York: Harcourt Brace & Co., 1952.

Hartman, R.R.K. and Stork, F.C. Dictionary of Language and Linguistics, London: Applied Science Pub., 1972.

Hutchins, W.J. Language of Indexing and Classification, Herts, England: Peter Peregrinus, 1975.

Jones, Spark Karen. Information Retrieval A Draft for Encyclopedia of Artificial Intelligence, 1990.

Katz, J.J. and Postal, P.P. An Integrated Theory of Linguistic Description, Cambridge: The MIT Press, 1964.

Lancaster, F.W. Information Retrieval System; Characteristic, Testing and Evaluation, New York: John Wiley and Sons Inc., 1968.

Langacker, Ronald W. Fundamentals of Linguistic Analysis, New York: Harcourt Brace Jovanuvich, 1972.

Lyons, John. Introduction to Theoretical Linguistics, Cambridge: At the University Press, 1968.

Neelameghan, A. "Research on the Structure and Development of the Universe of Subject", Library Science, V. 4, 1967.

Neelameghan, A. and Maitra, R. "Non-Hierarchical Associative Relationships in Social Sciences; Identification & Typology", DRTC Seminar, V. 15, 1977.

Paillet, Jean-Pierre & Dugas, A. Approaches to Syntax, Armsterdam: John Benjamin Pub. Co., 1982.

Richards, J. Longman Dictionary of Applied Linguistics, Essex, England: Longman, 1985.

Soergel, D. Organizing Information, Principles of Data Base and Retrieval Systems, London: Academic Press, 1985.

Vickery, B.C. On Retrieval System Theory, 2nd Edition, London: Butter Worths, 1965.

CHAPTER III

PRINCIPLES OF INFORMATION RETRIEVAL SYSTEM BASED ON LINGUISTIC ANALYSIS

PRINCIPLES OF INFORMATION RETRIEVAL SYSTEM BASED ON LINGUISTIC ANALYSIS

3.0. Introduction

After making conceptual study of linguistics and information retrieval system and their relationships, now we will try to trace out principles of information retrieval systems based on linguistics. As we have already stated in the earlier chapter that only two levels of linguistics are useful for the purpose of information retrieval. These are: syntax and semantics. Therefore, we will try to concentrate on these two levels for identifying principles propounded by different scholars of the field.

We are not going into the controversy of the technical terms used by different authors for the term 'principle'. Our main purpose here is to see how these normative principles, or guiding principles or principles are helpful in interpreting and explaining the empirical facts of experience found in the field of library and information science. Any term may be used to denote the idea. Coining technical terms is not our concern here.

Library and information science is a very vast discipline. It is quite difficult, if not impossible, to scan each and every thinker, who has contributed for the development of the discipline. For our purposes we will try to incorporate only those theoreticians who have contributed in the field of subject indexing and information retrieval. Here again, our concern is only to find out whether principles or normative principles propounded by these authors are based on linguistics with special reference to the levels of syntax and semantics.

Knowledge representation is the key to effective retrieval and dissemination of information. Knowledge is generally represented in a subject statement. A subject statement consists of various terms. Terms stand for concepts and concepts combine in the statement according to a pattern of relationships. Thus we can say that terms (vocabulary) and term ordr and their linking mechanism make an indexing language. In other words, indexing languages have two kinds of relationship: one operating at the level of the concept, and the other concerned with the combination of concepts. The first concerns those familial and other categorical relationships which link a particular concept to the class or classes of ideas of which it is a member, in this way, for example, that

the entity 'Dog' is a member of such categories as 'Canines', 'Carnivore' and mammal. The second concerns those syntactic links between the separate elements of a compound subject which endow a subject statement with meaning, in the way of example, that the terms 'Dog', 'Sheep' and 'Herding' can be taken together in a particular way to express the subject 'Herding of sheep by dog'. The first of these are present by implication and can be distinguished as a priori relationships, while the second cannot be known until after they have been made explicit, and so are a posteriori relationships. They can be recognised as Gardin's paradigmatic and syntagmatic relationships respectively, and also distinguished as: (a) semantic or categorical relationship; (b) syntactic relationships. (Austin 1976: 168)

Semantic Relationships

Semantic relationship is a familial relationship. In this, we try to search an alternative term or substitute term. Semantic relationships are categorised in three groups: Equivalence, hierarchical, and affinitive/ associative.

(a) Equivalent: Semantic relationships include:

1. Synonyms and antonyms
2. Quasi synonyms

3. Same continuum
4. Overlapping
5. Preferred spelling
6. Acronyms, abbreviations
7. Current and established terms
8. Translation
9. Technical terms and layman terms.

(b) **Hierarchical: Semantic Relations include:**

1. Genus - Species
2. Whole - Part

(c) **Affinitive/Associative: Semantic Relations include:**

1. Co-ordination
2. Genetic
3. Concurrent
4. Cause and effect
5. Instruments
6. Materials
7. Similarity
8. Concomitant terms
9. Cow-calf principles.

We can show the semantic relationship in the field of information retrieval by devices like cross-reference, by using symbols like NT (Narrower term),

BT (Broader term) and RT (Related term). This allows the reader to broaden or narrow the scope of his search, and thus help to improve the recall or precision according to his/her need. Even in the classification schemes, these relationship of hierarchical, affinitive, and associative can be shown by grouping together related concepts in a systematic arrangement. Semantic relationship is very useful in providing alternative or substitute term at the search stage and thus facilitates the recall value.

Syntactic Relationships

Syntactic device in indexing languages are of two types: word order, and relator or linking mechanism. Derek Austin puts it as, "subject statement concerns (a) kinds of concepts, and (b) relationship between concepts. Rules for ordering the parts of a compound subject into a linear sequence could theoretically be based on either of these factors" (Derek Austin 1976: 181). The primary purpose of term order is arrangement. It is true that order usually determines meanings. According to Needham, the use of word or term order and relators as syntactic devices are common in indexing languages:

> Ranganathan's PMEST, Coates's Thing - Material - Action and relationship table. Vickery's 'standard' citation order, are some of the examples of formulae for determining term order.... These can be regarded

> as attempts at standardizing and controlling syntax - as were earlier rules by Cutter, the Thing - Process sequence of Kaiser, the grammatical approach of Prevost, etc. (Needham 1971: 247)

Now we will try to examine contributions of different scholars. As stated earlier, we will restrict ourselves to those thinkers, who have contributed their theories based on linguistics in general and its levels: semantics and syntactics in particular.

3.1. Principles Propounded by Different Scholars

3.11. Charles Ammi Cutter

The first attempt to systematic study of subject indexing was made by Charles Ammi Cutter. His book, Rules for a Dictionary Catalogue, which appeared in 1876 may be regarded as pioneer work in subject indexing. In this book, he propounded several principles on syntax and semantics for subject indexing. Broadly, his main principles can be categorized in three parts:

(i) Principles relating to specific entries.

(ii) Principles of usage of entries.

(iii) Principles of syndetic structure.

Out of these three categories of principle, (i) and (iii) are pertaining to semantics whereas (ii) relates to syntax.

(i) **Principles of Specific Entries**

Cutter's principle of specific entry is fundamental. This principle makes provision that the heading should be co-extensive with (no broader or narrower than) the subject content of the work. According to Cutter, "Enter a work under its subject heading, not under the heading of a class, which includes that subject" (Cutter 1876: Rule 161). For example, a book on "The Cat" should be entered under 'CAT' and not under "Zoology" or 'MAMMALS' or 'DOMESTIC ANIMALS'.

The characteristics of specific entry of Cutter is as follows:

(1) Enter a work under the heading which expresses its subject precisely.
(2) If a subject cannot be expressed in a single word or customary phrase consisting of adjective followed by a noun, are to be specified exactly by a compound heading consiting of more than one word.
(3) In such cases entry word will be a class-term relative to the specific subject.
(4) It will be followed by the further term as sub-heading.

(5) The meaning of headings and sub-heading in combination must be clear and self explanatory.

The following examples illustrate these in practice:

Flats, Architecture
Floods, Social Relief
Distillation, Chemical Technology
Marriage, Psychology
Algae, Ecology.

These examples help to clarify the meaning of specific subjects. Cutter claims that "the rule of "specific entries" is the main distinction between the dictionary catalogue and the alphabetico-classed catalogue" (Cutter 1876: 67).

(ii) Principle of Usage of Entries

Cutter is of the view that in the compound heading of any kind, raises the important question of the order in which constituent parts are to be cited. In other words, which term in effect, is to be the all important entry word. Rules 174 and 175 describe how the constituent words of phrases are to be ordered in the heading. In

Rule No. 174, he points out types of compound subject heading as:

The name of subject may be:

(a) A single word, as Ethics, Economics, Botany or several words taken together.
(b) A noun preceded by an adjective, as Capital punishment, Ancient history.
(c) A noun preceded by another noun used like an adjective, as Death penalty, Flower fertilization.
(d) A noun connected with another by a preposition, as Penalty of Death, Fertilizer of flowers.
(e) A noun connected with other by 'and', as Ancient and Modern.
(f) A phrase or a sentence.

For above compound headings, he suggests three types of phrases that can be inverted.

(a) The noun preceded by an adjective, such as Agricultural Chemistry;
(b) The noun preceded by an adjectival noun such as Death penalty; and

(c) The prepositional phrase such as Fertilization of Flower.

He also discusses three possible rules for these:

(i) Never invert;

(ii) Invert only type (a); and

(iii) Invert any type of compound name, if doing so, brings most significant word of the phrase.

Cutter suggests in his Rule No. 175:

> Enter a compound subject name by its first word, inverting the phrase only when some other word is decidedly more significant or is often used alone with the same meaning as the whole. (Cutter 1876: Rule No. 175)

For examplee:

(A) Death penalty to DEATH PENALTY.

(B) Agricultural Chemistry to CHEMISTRY AGRICULTURE.

(C) Natural Science to PHYSICS.

(iii) Principles of Syndetic Structure

Cutter suggests that the catalogue should be syndetic, i.e., it should include a system of references relating to the subject scattered by alphabetical arrange-

ment. These references serve two purposes: (1) to guide users from their search terms to valid headings; and (ii) to link semantically related headings, i.e., equivalence, heirarchical and associative. The following rules propounded by Cutter are evidence of these semantic relations.

Rule 168 Of two exactly synonymous names choose one and make a reference from the other, as Poison and Toxicology.

Rule 169 In choosing between synonymous headings prefer the one that:

(a) is most familiar to the class of people who consult the library;

(b) is most used in other catalogues;

(c) has fewest meanings other than the sense in which it is to be employed;

(d) comes first in the alphabet, so that the reference from the other can be made to the exact page of catalog; and

(e) brings the subject into the neighbourhood of other related subjects.

Rule 171 Of two subjects exactly opposite, choose one and refer from the other.

Rule 187 Make references from general subjects to their various subordinate subjects and also to coordinate and illustrative subjects.

Rule 188 Make references occasionally from specific to general subjects.

If we look these rules we find that Rules 168, 169 and 171 deal with equivalence relation whereas Rules 187 and 188 deal with hierarchical and associative subjects.

Rules of Dictionary Catalogue though overtaken by more recent theories, still form the basis of much practice and for this reason - and because it cannot be said that all the problems Cutter dealt with have been solved - they are still important.

Cutter's Rules remain the fountain head of all later American approaches to information both descriptive cataloguing and subject analysis. His Rules are the basis for the Library of Congress List of Subject Headings (LCSH) and its progeny; the Sear's List of Subject Headings.

3.12. Julius Otto Kaiser

Cutter has been often criticised on account of

his ideas in respect of 'citation order' or 'significant order' among the components of a compound subject. These problems were taken up by Julius Otto Kaiser. He published a book named Systematic Indexing in 1911, which is a good systematic approach to an alphabetical index to information. His main contributions in the field of Subject Indexing, which are related to syntax and semantics, are as follows:

1. He named subject expressions as statements.
2. The terms in a statement fall into two categories, viz., Concretes and Processes. (Kaiser 1911: Para 298).

   (a) Concrete implies a substance or a concrete entity. They can be divided into:
      (i) Movable - silk, hardware etc.
      (ii) Immovable - land, rivers, resources, etc.
      (iii) Abstract - labour, mental, manual etc. (Kaiser 1911: Para 299)

   (b) Process represents the condition of concretes, what they do or what is done to them? Feeding, cultivating etc. (Kaiser 1911: Para 301)

(c) Immovable concretes include one another kind of special importance i.e., countries in political sense. We are obliged to treat the political divisions called countries as a distinct class. We now have:

Concrete { Movable / Immovable, except countries / Abstract

Countries. (Kaiser 1911: Para 300)

Three possible combinations of these terms with examples are:

(i) Concrete and process ... Wool - scouring.

(ii) Country and process ... Brazil - Education.

(iii) Concretes, Country and process ... Nitrate - Chile - Trade.

3. The categories in a statement are to be arranged in the sequence of:

(i) Concrete - Country - Process, or

(ii) Country - Concrete - Process.

4. No statement is to be made under process.

The semantic part of Kaiser's indexing language to express cross references is measured with the help of:

(i) Higher collectives

(ii) Lower specifics

(iii) Synonyms

For example,
Stone < Building material
> Granite, sandstone, marble
= Wood, brick, cement

Thus, we find that the system of Kaiser has its own semantics and syntax and can be considered as precursor to two of the most significant developments in the 20th century, namely:

(i) Emergence of concept of indexing language; and

(ii) The use of facet analysis to categorise subject indicators which was, later on, followed by Ranganathan in his 'colon classification'.

3.13. James Duff Brown

J. D. Brown being dissatisfied with Dewey system of its American bias, designed a British system of classification known as 'Subject Classification Scheme' which appeared in 1906 and the latest edition, the 3rd (edited by J.D. Stewart) in 1939. Distinctive features of Brown's classification scheme are as follows:

1. One Place Theory

Brown points out that the variety and intersection of the departments of knowledge is so great that it is not possible to show the complete literature of a special subject at one fixed place.

> Every subject is capable of being treated from a large number of standpoints, and each of these may be the centre of enormous literature ...; e.g., the Subject 'Rose' may be viewed from the standpoint of Botany, Horticulture, History, Geography, Decoration, Symbolism, Bibliography etc. (Brown 1906: 8-10)

He considers 'Rose' to represent a concrete subject, whilst the standpoints represent general subjects. "All the concerete topics are given only one place in the schedule. The standpoints which qualify them are listed separately in a categorical table of forms, phases, standpoints, qualifications etc." (Brown 1906: 15). From above discussion it is clear that:

(a) The main idea of SC is that the subjects are 'concretes'.

(b) The subdivisions of a subject are aspects of these concretes and are consequently treated in the categorical tables.

2. Science and Application Theory

Every class is arranged in a systematic order of 'scientific progression', as far as it seemed possible ...; while applications directly derived from a scientific or other theoretical base, have been placed with that science or base. (Brown 1906: 11). For example, Rose under Botany, Coal under Mineralogy.

3. Facet Analysis in the SC

The principle of providing Categorical Tables certain concepts, which apply to many subjects automatically leads to a certain degree of facet analysis, e.g.:

(a)
A 900 General Science
.10 History
A 900.10 History of Science

(b)
I 220 Gardening
R 000 France
I 220 RO Gardening in France.

4. Arrangement Order

Main class order is based on Matter, Life, Mind and Record. The arrangement order is broadly based on the following caterogies:

(i) The 'concrete' is preferred to the 'general' subjects;

(ii) Applications follow their theoretical base; and

(iii) The place of constant interest is preferred to the place of occasional interest, e.g., Library Architecture under Library Economy, not Architecture.

One place theory and the principles of Science and Application Theory reveal semantic relationship. The main idea of Brown – 'Concrete' is influenced by the ideas of Cutter's 'specific entry', which leads toward specialization and is based on semantic relation Genus and Species. Again, the principle of Science and Application shows same semantic relationship.

Facet analysis in Categorical Table with concrete subject and the citation order shows syntax in the indexing language.

The lack of success attending this scheme should not preclude its study. It contains germs of several ideas that are now being discovered to be fundamental principles. Kaiser & Coates have used the concept of 'concrete' in alphabetical subject indexing. The approaches

of Barbara Kyle & Farradane have been similar to Brown. Theoreticians like Ranganathan worked on the concept of concrete to give the 'Postulate of Concreteness'.

3.14. Henry Evelyn Bliss

Bliss is a formidable theorist in the field of subject indexing whom we cannot ignore. His main contributions are outlined in his two volumes: The Organization of Knowledge and the System of Sciences (1929), and Organization of Knowledge in Libraries and the Subject Approach to Books (1933: 2nd ed. 1939), and subsequently in the introduction of his classification scheme namely, Bibliographic Classification (1940-1953) in four volumes.

We find from the beginning that several principles guided Bliss's work. These are consensus, collocation of related subjects, subordination of special to general, gradation in speciality, and the idea of alternative location. All these principles pertain to semantic needs of an indexing language.

As far as consensus is concerned, Bliss is of the opinion that:

> Knowledge should be organized in consistency with the 'scientific' and 'educational consensus', which is relatively stable and tends to become more so as

> theory and system become more definitely and permanently established in general and increasingly in detail. (Bliss 1939: 42-43)

He advocated that such as order would be very helpful to the users of the library.

The original classification of Bliss was an 'aspect' classification or classification by discipline in which information on a particular 'phenomena' was scattered over many disciplines and sub-disciplines. For example, Iron was subordinated to many subjects like Chemistry, Chemical technology, Mineralogy, Mining, Industrial economics etc. Later on, influenced by the idea of Brown, he tried to collocate the related subjects by bringing them into close proximity. This was achieved by formulating three principles, e.g., subordination of special to general – in bringing special subjects under comprehensive general subjects. Gradation in speciality and Alternative Location. Whereas the principles of subordination of special to general has been followed by Bliss in forming chains, i.e., to bring special subjects under comprehensive general subjects, the principle of gradation in speciality has been used to arrange coordinate topics.

The basic idea of this principle is that certain derivative subjects draw on the findings of other subjects or in other words, there are certain subjects which depend upon the finding of other subjects. In this respect, those which are dependent become more specialized than the subjects from which ideas are borrowed. According to this principle, dependent subject should follow (often coordinate) topics on the finding of which it has to rely upon. It appears that perhaps Bliss had been influenced in this respect by the principles of serial dependence, formulated by Auguste Cumte.

Further to make the systems more useful from the point of view of user, Bliss suggested the principles of 'Alternative Location'. According to this principle, a subject may be placed with equal usefulness in two or more possible locations in the scheme. For example, Economic history can be subordinated to general history or classed in Economics.

New edition of 'BC' which is greatly influenced by Ranganathan's idea of faceted classification and citation order shows many features added to the scheme, specially in its syntactical structure. Here are some examples which prove this claim.

The principles and method of facet analysis and structure is central to the BC2, which is fully analytico synthetic scheme. For example:

P. = Religion.

POZ = Religion in individual countries.

POZ RB = Religion in China.

BC2 is a fully synthetic scheme, with common schedules, special schedules, introduction of divide like other parts of scheme and retrospective number building.

BC2 recognizes Phase relations propounded by Ranganathan. It does recognize general relation, comparison, bias, influence. Ranganathan's difference phase relation is not in BC2.

It is clear from above discussion that BC2 is greatly influenced with the syntactical ideas of Ranganathan.

3.15. Shiyali Ramamrita Ranganathan

Ranganathan is widely known for his contributions to the field of Library and Information Science. His contributions are not only vast and varied but also original. As a librarian, teacher, Chairman in different

national and international committees like UNESCO, FID etc., he blended his vast and rich experience in his monumental work touching every aspect of Library and Information Science. But his outstanding contributions are in the field of Information Retrieval and specially in Classification.

R. K. Olding has rightly remarked that:

> There is hardly a single aspect of the librarianship to which Dr. Ranganathan has not turned his attention, but the field in which his greatest influence has been felt is that of Library Classification. A whole school of thought has been developed embodying the principlees of classification first laid down by him. (Olding 1967: 189).

The whole theory of information retrieval propounded by Ranganathan is based on a conceptual framework of Laws, Canons, Postulates and Principles.

Ranganathan has also prescribed hierarchy of these guiding ideas. This is as follows:

(1) Basic Laws

(2) Fundamental Laws

(3) Canons for Classification

(4) Principles for Helpful Sequence

(5) Postulates for Classification; and

(6) Principles for Facet Sequence.

(Ranganathan 1967: 113)

Above hierarchy of guiding ideas (Nos. 3-6) can be divided into two parts: Semantics and Syntax of indexing language. Canons for work in the idea plane: Canons for Characteristics, Canons for Array, Canons for Chain, Canons for Filiating Sequence, Canons for work in the verbal plan and Principles for Helpful Sequence guide the semantic part of the indexing language, whereas Postulates and Principles for Facet Sequence deal with the syntactic relationships of indexing language. A detailed list of these Canons, Postulates and Principles is given below:

Syntax

Postulates for Classification

(A) 1 Postulate of Fundamental Categories

Postulate of Basic Facet

Postulate of Isolate Facet

Postulate of Rounds for Energy

Postulate of Rounds for Personality and Matter

Postulate of Round for Space and Time

Postulate of Level

Postulate of Facet Sequence

Postulate of First Facet

Postulate of Concreteness

Postulate of Facet Sequence Within a Round

Postulate of Facet Sequence within Last Round

Postulate of Level Cluster.

(B) Principles for Facet Sequence

1. Wall-Picture Principle
2. Whole-Organ Principle
3. Cow-Calf Principle
4. Actand-Action-Actor-Tool Principle

Semantics

(A) Canons for Work in the Idea Plane

(1) Canons for Characteristics

Canon of Differentiation

Canon of Relevance

Canon of Ascertainability

Canon of Permanence

Canon of Concomitance

Canon of Reelevant Succession

Canon of Consistant Succession

(2) Canons for Array

Canon of Exhaustiveness

Canon of Exclusiveness

Canon of Helpful Sequence

Canon of Consistent Sequence

(3) **Canons for Chain**

Canon of Decreasing Extension

Canon of Modulation

(4) **Canons for Filiatory Sequence**

Canon of Subordinate Classes

Canon of Coordinate Classes

(5) **Principles for Helpful Sequence**

Principle of Later-in-Time

Principle of Later-in-Evolution

Principle of Spatial Contiguity

Principle of Bottom-Upwards

Principle of Top-Downwards

Principle of Left to Right

Principle of Right to Left

Principle of Clockwise Direction

Principle of Counter Clockwise Direction

Principle of Periphery to Centre

Principle of Centre to Periphery

Principle of Away-from-Position

Principle of Increasing Quantity

Principle of Decreasing Quantity

Principle of Increasing Complexity

Principle of Canonical Sequence

Principle of Literary Warrant

Principle of Alphabetical Sequence

(B) **Canons for Work in the Verbal Plane**

Canon of Context

Canon of Enumeration

Canon of Currency

Canon of Reticence

Syntax

It is found convenient to divide universe of subjects as a first step, into a few mutually exclusive and totally exhaustive set of Main Subjects. These are not too many in number. Therefore, most of the schemes succeed in managing them on the basis of some principles. The resulting sequence more or less helpful in most cases. There is not much to choose between them. Therefore, we may take the main subject and their sequence to be postulated by the scheme prescribing them.

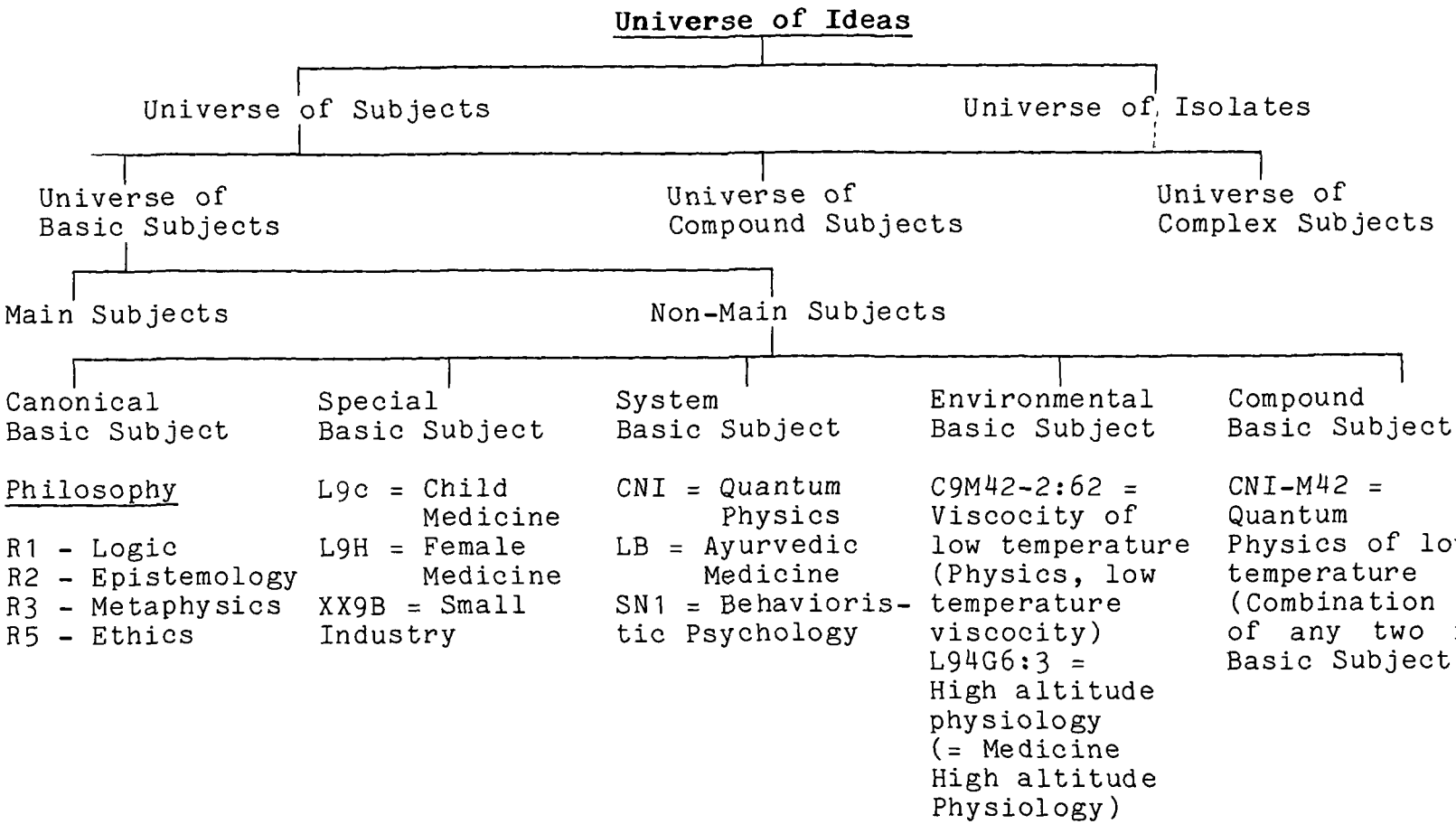
"A subject is made of a Basic Subject only or it has a Basic Subject and one or more isolate ideas as its components. In the latter case, it is a Compound Subject" (Ranganathan 1967: 83). For example, Mathematics,

Physics, Chemistry, these are all Basic Subjects, whereas subject like 'Indian Law' or 'Law of Crime', Geometry of Solids, are Compound Subjects.

Isolate ideas on its own, cannot be a subject. Ranganathan defines them as, "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" (Ranganathan 1967: 83). For example, child it is a part of a subject as child medicine, but not fit to be a subject.

Now a question arises, what is Basic Subject? "A Basic Subject without any isolate idea as a component is a Basic Subject" (Ranganathan 1967: 83). For example, Geometry, Physics, Chemistry are all Basic Subject.

The first Postulate of Ranganathan specifies that every subject will have a Basic Subject. A Basic Subject may be a Main Subject or a Non-Main Subject. The following is a schematic representation of different kinds of Basic Subjects:


Universe of Ideas
Universe of Subjects
Universe of Isolates
Universe of Basic Subjects
Universe of Compound Subjects
Universe of Complex Subjects
Main Subjects
Non-Main Subjects
Canonical Basic Subject
Philosophy
R1 - Logic
R2 - Epistemology
R3 - Metaphysics
R5 - Ethics
Special Basic Subject
L9c = Child Medicine
L9H = Female Medicine
XX9B = Small Industry
System Basic Subject
CNI = Quantum Physics
LB = Ayurvedic Medicine
SN1 = Behavioristic Psychology
Environmental Basic Subject
C9M42-2:62 = Viscocity of low temperature (Physics, low temperature viscocity)
L94G6:3 = High altitude physiology (= Medicine High altitude Physiology)
Compound Basic Subject
CNI-M42 = Quantum Physics of low temperature (Combination of any two non-Basic Subject)

Further, to meet the challenges of complicated phenomenal level of isolate ideas and the difficulty of choosing a particular succession of immediate neighbourhood relation of remove 1, 2 and so on; and securing the same pattern of arrangement in all the homogeneous sets of Compound Subjects among the very large number of subjects, irrespective of basic subject, Ranganathan developed General Theory of Classification. It suggests a deep drive to a near-seminal level to reach something practicable. Work in the near-seminal level unveiled the possibility of reducing to five types of large variety of the isolate facets presented by all the subjects of the Universe of Subjects taken as a whole. This gave rise to the postulate of Fundamental Categories. According to it, "There are five and only five Fundamental Categories; viz., Personality, Matter, Energy, Space and Time" (Ranganathan 1965: 65).

The significance of these categories lies in the statements about the facets of the subject – their separation and sequence. For the brevity sake, these are denoted by the intionym PMEST. These categories are as follows:

Time

Any concept representing a period of year, month, season, etc. is a manifestation of time. Generally, this category is measured by the Christian era. In addition to this, it includes featured time – such as, day, sunrise, noon, afternoon, evening, night, mid-night, twilight, seasons and meteorological periods.

Space

The surface of the earth, the space inside it and the space outside it, are manifestations of the Fundamental Cetegory of 'Space'. Usually it contains Geographical isolate ideas - such as continents, countries and counties - and water formations - such as oceans and seas are taken to be its manifestation. Physiographical Isolate Ideas - such as desert, prarie, rainforest, plateau, mountain, river and lake are also taken to be manifestation of the Fundamental Category 'Space'. So also an area of population-cluster - such as a city, a town, and a village is taken to be the manifestation of the fundamental category 'Space'.

Energy

Any external action or operation is a manifestation of the fundamental category 'Energy'. For example,

'measuring', 'teaching', 'analysing', 'classifying' all are the manifestations of the FC 'Energy'. A manifestation of energy will always involve a 'method' of doing the 'action', an 'agent' of action (and often a tool).

Matter

Its manifestations are taken to be of two kinds – Material and Property. Matter can be defined as any attribute. Generally it includes:

(i) Intransitive verb situations, e.g. 'learning', 'digestion' in the contexts
Child; learning
Man; digestion etc.
These can be referred as self activities;

(ii) Concepts denoting functions of an entity such as 'Physiology' in Human body; Physiology, etc.;

(iii) Abstract properties or assessed properties such as beauty, efficiency, effectiveness, intelligence, etc.

(iv) Intrinsic properties such as density, thickness, intrinsic constituent in
Coffee; Caffeine content
Bear; Alcohol content
Metal; Density, etc.

Personality

Personality is to be defined as the core concept in a compound subject. For example,

Social Sciences - community

Chemistry - elements and compounds

Life Sciences - living systems and their organs,

etc., are all manifestations of the Fundamental Cetegory 'Personality'. In every subject field, it is possible to identify a core concept constituting the core entity of study of the concerned subject field.

After postulating the Five Fundamental Categories, Ranganathan felt the need for more Postulates to explain the re-occurrence of these categories in the same subject. For this purpose he developed two new concepts. These are: Rounds and Levels. The postulates governing Rounds and Levels are as follows:

Postulate of Rounds for Energy

"Energy may manifest itself in one and the same subject more than once, that is, more than one Round" (Ranganathan 1965: 67). For example:

(1) Diagnosis of disease = L:4:3 [CC ed 6]

L = Medicine (BS)

4 = Disease [E]

3 = Diagnosis [2E]

(2) Ayurvedic cure = LB:4:6 [CC ed 6]

L = Medicine (BS)

LB = System of Medicine (System Basic Class) [P1]

4 = Disease [E]

6 = Cure [2E]

Postulate of Rounds for Personality and Matter

It is possible for a manifestation of 'Personality' and 'Matter' to occur after [1E] again after [2E] again after [3E] and so on: that is, in any Round. (Ranganathan 1965: 68). For example:

(i) Tuberculosis in poultry = KZ7351:421

KZ Animal Husbandry (BS)

KZ7351 Poultry [P]

KZ7351:4 Disease [E]

KZ7351:421 Tuberculosis [E2P]

(ii) Control of Cholera = L25:425:55

L = Medicine (BS)

25 = Intestine [P1]

4 = Disease [E]

25 = Metaphyta [2P]

5 = Prevention [25]

55 = Control Measure [3P]

Postulate of Level

"Personality may manifest itself in one and the same Round in a subject more than once; i.e., in two or more levels, so also in the case of Matter, Space and Time" (Ranganathan 1965: 68). For example:

Sakuntala of Kalidas = 015,2D40,3

0 = Literature (BS)

15 = Sanskrit Literature [P1]

2 = Drama [P2]

D40 = Kalidas [P3]

3 = Sakuntala [P4]

Postulate of Level Cluster

"Facet of different Levels of the same Fundamental Category within a Round should be kept together" (Ranganathan 1965: 69)

Postulate for Space and Time

"Ordinarily [S] and [T] should be put last in the sequence in which they are mentioned here" (Ranganathan 1965: 69). For example:

Basic School in India in 1950s = TN3.44'N5

T = Education (BS)

TN3 = Basic School [P1]

44 = India

N5 = 1950s

To arrange these Fundamental Categories, Ranganathan postulated the sequence of these categories in the order of decreasing concreteness. This postulate reads as, "The Five Fundamental Categories fall into the following sequence when arranged according to their decreasing concreteness: [P], [M], [E], [S], [T]" (Ranganathan 1965: 68-9). For example:

Selection of periodicals in University library in India during 1995 = 234;46:1.44'N95

2 = Library Science (BS)

34 = University Library [P1]

46 = Periodicals [1M1]

1 = Selection [E]

44 = India [S]

N95 = 1995 [T]

Principle for Facet Sequence

Apart from these Ranganathan has enunciated four 'Principles of Facet Sequence'. These act as guidelines

for determining the sequence of Facets, in a subject. Out of these four 'Principles for Facet Sequence' the Wall-Picture Principle is the basic one, the others are derivatives of it. The Wall-Picture Principle guides as:

> If two facets 'A' and 'B' of a subject are such that the concept behind 'B' will not be operative unless the concept 'A' is conceded, even as a mural picture is not possible unless the wall exists to draw upon, then the facet 'A' should precede the facet 'B'". (Ranganathan 1967: 425)

For example, in the title 'Treatment of disease of human ear' the concept behind 'Disease' will not become operative unless the concept behind "Ear" is conceded. Therefore, these two Isolate Facet should be put in the sequence Ear, Disease. Similar is the case with the concepts of 'Treatment' and 'Disease'. Therefore, the sequence of these Isolates will be Disease, Treatment. Thus the sequence of the Basic Facet and Isolate Facets in the Subject will be:

Medicine, Ear, Disease, Treatment = L183:4:6

[CC ed 6]

L = Medicine (BS)

183 = Ear [1P1]

4 = Disease [1E]

6 = Treatment [2E]

Other principles are derivatives of the Wall-Picture Principle. These include:

(1) Whole-Organ Principle;

(2) Cow-Calf Principle; and

(3) Actand-Action-Actor-Tool Principle.

On the basis of the postulates mentioned earlier a General Model of Facet Sequence or Syntax has been proposed by Ranganathan. He has named it as "Generalized Facet Structure". It reads as follows;

(BS)

[1P1], [1P2], ...,[1PLp]

[1M1]; [1M2]; ...; [1Mlm]

[1E]:

[2P1], [2P2], ..., [2PLq]

[2M1]; [2M2]; ...; [2MLn]

[2E]:

... ...

[RP1], [RP2], ..., [RPLr]

[RM1]; [RM2]; ...; [RMLr]

[RE]:

[S1]. [S2]. ... .[SLs]

[T1]' [T2]' ... '[TLt]

Where Lp, Lm, Ln, Lr, Ls, Lt may have any integral value. (Ranganathan 1967: 464).

Semantics

Coming to the semantics propounded by Ranganathan, we find that Canons for characteristics, Canons for Array, Canons for Chain, Canons for Filiatory Sequence, Canons for Verbal Plane and Principles for Isolate Sequence deal with semantic relationship in one way or the other.

The problems of Verbal Plane concern naming each of the isolates in the schedule of isolates contributing to the same kind of facet in all the Compound Subjects of a related set. Convenience of use requires standard terminology for naming Basic Subject and Isolate Ideas. These two groups constitute the fundamental constituent of each of the Compound terms. To meet these requirements, standard glossaries have been compiled in different countries.

Ranganathan is of the opinion that standard terminology should be used for every concept or term. The terms used in indexing language should be free from

homonyms and synonyms. To resolve this, Ranganathan formulated the Canon of Context which says that:

> The denomination of a term in a scheme for classification should be determined in the light of the different classes, or ranked isolates of the lower order (upper link) belonging to the same primary chain as the class, or the ranked isolate, denoted by the term in question. (Ranganathan 1967: 208)

However, the denotation of a term may not be the same to different persons. Uniformity cannot be imposed. The solution to this problem lies in the Canon of Enumeration which reads as follows:

> The denotation of a term in a scheme for classification should be determined and should be left to be determined in the light of or through the sub-classes or rankeed Isolates (lower links) enumerated in the various chains having the class or ranked isolate, as the case may be, denoted by the term in question as their common link. (Ranganathan 1967: 211)

The Canons of Classification for work in the Idea Plane

The Canons of Classification for work in the Idea Plane fall into four sets:

1. Canons for Characteristics (7 Nos)	15 Nos
2. Canons for Array (4 Nos)	
3. Canons for Chain (2 Nos)	
4. Canons for Filiatory Sequence (2 Nos)	

These Canons of Classification for the Idea Plane belong to semantic level as they provide guidelines for hierarchical relationships.

The Canons for Characteristics help in sorting the isolates of a universe into sub-aggregates, whereas the Canons for Chain, Canons for Filiatory Sequence, Canons for Array and the Principles of Help Sequence guide hierarchical relationships.

The Canon of Decreasing Extension prescribed that:

> While moving down a chain from its first link to its last, the extension of the classes, or of the ranked isolates, as the case may be, should decrease, and the intention should increase at each step. (Ranganathan 1967: 174)

For example:

World
|
Asia
|
India
|
Meghalaya
|
Shillong

Similarly, the Canon of Modulation also serves as a guide to hierarchical relationships in a chain.

It emphasizes on the maintenance of continuity without break of any relevant link between the first link and the last link.

Canons for the Filiatory Sequence require that the subordinate and coordinate classes be arranged in their degree of filiation. Canons for Array provide help as guidelines in formulating arrays of isolate ideas.

The 'Principles of Helpful Sequence' help in the determination of sequence among isolate ideas in an array. These principles have already been mentioned on page No. 76-77.

3.16. Gerard Cordonnier

About 1940s Cordonnier proposed to launch a 'Universal Classification Scheme'. While lecturing to UFOD in 1943, he gave outline of his classification schemes which was published in 25 divisions. He also stressed:

> The need to provide what he termed 'Functional endings', furnishing at least an elementary grammar ... to distinguish nouns, adjectives, verbs, subjects and direct or indirect objects or compliments and added these functional indicatives will from the object of special division of the classification, relating to practices of relationships. (Cordonnier 1944: 10)

A number of French specialised documentation organisations were greatly impressed and influenced by his classification ideas and consequently adopted his universal classification scheme for their organizations. However, Cordonnier, is relevant for us for providing a series of categories used in designing a special classification scheme for the Centre de Documentation des Constructions et Arms Navales. These categories are as follows:

"1. Organisms and Services (Origin, or Subject, of documents) (nature; places).
2. Persons (miscellaneous categories ...)
3. Individuals (living beings ...); biological conditions ...
4. Bodies (natural; simple, compound, ...) (miscellaneous condition)
5. Miscellaneous equipments (property, fittings...)
6. Miscellaneous actions (physical; technical; economic; ...)
7. Intellectual concepts
8. Documentary forms ...
9. Time ...." (Cordonnier 1960; 12-31)

Analysis of his work reveals that he has provided proper place for semantic and syntax in his classification. His ideas relating to 'particles of relationship' and concept of facet structure illustrate this fact.

3.17. Jasua Edward Louis Farradane

Farradane, a prominent member of CRG selected completely new area of his research – the analysis of relationships between concepts and developed his now well known "relational operators" (Farradane 1961: 127). These relators were based upon a theory in the psychology of thinking and they could also be used as the basis of facet analysis, although, Farradane's original idea was that they should be used for linking natural language terms in an alphabetical system.

The contributions of Farradane prominently lie in the dimension of syntax of an indexing language.

Farradane introduced three terms in his classification scheme – 'isolate', 'operator' and 'analet'. An isolate is a term that can be the name of anything or action, i.e., Sucrose, milk, bird, migration, cutting etc. An 'operator' is a relation between any two isolates and there are nine of them, based on psychology of thinking. An 'analet' is a subject analysis of a given document in terms of isolate and operators.

Farradane's theory is based on the way the mind learns as how to know natural phenomena through two mental processes:

(i) The time basis of the associations between phenomena, that is fixed or temporary; and

(ii) The similarities or distinctness between phenomena.

From the coordination of these two bases, we can derive four operators, and by adding the category of 'Neither' we can increase this to nine. Each operator is given a symbol found in a standard typewriters key-board.

Farradane started with the theory that there was no need of Basic Subject. Compound Subjects are to be constructed from the universe of concepts without referring to Basic Subjects. He recognised the following four concepts basing on psychological theory of learning process:

(i) Entities;

(ii) Activities;

(iii) Abstracts; and

(iv) Properties.

Farradane's another contribution lies in recognizing nine types of relationships among the concepts constituting subjects. He distinguished these nine relationships by correlating two series of characteristics: the first concerned with the 'temporal' nature of relationship which can be non-temporal (non-time), temporary or fixed (permanent); the second concerned with the distinctness with which the relationship is perceived. From the second point of view the relationship can be simply one of co-existence (concurrent), the two concepts being present together in consciousness without their mutual relation being otherwise defined than by this co-existence; it can be non-distinct the two concepts being in this case directly tied one to the another, but without their method of attachment being known precisely; finally

Farradane's Operators

	Non-Time Relation	Temporary	Fixed
CONCURRENT	Concurrent /θ	Comparison & Self activity /*	Association /;
NON-DISTINCT	Equivalence /=	Dimensional & State /+	Appurtenance /C
DISTINCT	Distinctness /⊃	Reaction /-	Causation or Functional Dependence /:

the relationship may be distinct, the two concepts then being clear and explicit, but still definitely related. (Grolier 1962: 94).

These two series of characteristics can be used to form a matrix as shown with figure above. The points of intersection denote nine different kinds of relationship. These are as follows:

Concurrence

In this, we find mere existence of two terms in a subject, which can be found in language by the use of the word 'and' – A and B, for example,

Chemistry and Encyclopaedia

Chemistry /⊖ Encyclopaedia

Comparison

It is self expressive. A in comparison with B. It is also applicable to self activity, i.e., Man /* walking.

Association

It expresses the relation meant by the word 'for' – tool, for the job or process i.e., Government /; Library.

Equivalemce

This expresses identity or some degree. It gives

the meaning of 'as', i.e., Molasses (used) as Fodder or Molasses /= Fodder.

Dimensional

It implies the normal relation of time and place, i.e., in the morning on the table at Paris, in 1900. It also expresses 'States' or the temporary conditions of e.g., temperature etc. Ball /+ Box.

Appurtenance

(i) It expresses physical properties, e.g., Lead /C melting.

(ii) Part of something or permanent constituent, e.g. Bicycle /C Brake.

Distinctness

This expresses the realization of differences, considered alone. It can be used for the idea of imitation or substitution, e.g., Diamond / Glass.

Reaction

It expresses the action of one thing upon another or of a process upon a thing, e.g., Sugar /- Acid.

Causation or Functional Dependence

It implies the relation B arises out of A, or A causes B, but it should express only a direct cause and effect, e.g. Ore /: Iron, Author /: Book.

3.18. Hans Peter Luhn

Significant word has been a perennial problem in the retrieval and dissemination of information while confronting with a title. This problem was faced due to inexpressive titles of the works or due to semantic problems. To eliminate these problems in the selection of 'significant word', many efforts had been made in the past by several distinguished scholars like Andrea Crestadoro, Charles C. Jewett, and C.A. Cutter. Peter Hans Luhn also following his predecessor's idea of significant word tried to eliminate these problems with the help of computer. He gave new name to the above idea and developed a device as 'Keyword - in context' (KWIC) in indexing during 1950s.

In 'keyword-in context' indexing system, 'keywords' or significant words are extracted from the title of a document. The title of the document is manipulated in such a manner that each significant word becomes entry element. This is done with the help of computer on the basis of Linguistic idea of "Content words and functional words." By content words we mean 'Words which have statable LEXICAL meaning in fact, apart from the few FUNCTIONAL words', whose role is primarily to express GRAMMATICAL relationships." (Crystal 1980: 86). The process of extract-

ing significant words (content words) depends on some rules to differentiate what is significant and non-significant. Since significant word is difficult to predict, it is more practical to isolate it by rejecting all obviously non-significant or common words, with the risk of admitting certain words of questionable status. Such words may subsequently be eliminated or tolerated as so much 'noise'. A list of non-significant words would include conjunctions, prepositions, auxiliary verbs, certain adjectives and words like 'articles', 'reports', 'analysis', 'theory' etc. The remaining significant words or keywords would be extracted from the title together with a certain number of words that precede and follow them. By arranging these words in preceding and following positions in alphabetical order of the keywords, the KWIC index is generated. (Schultz 1968: 229). For example, the following titles:

(i) The binomial theorem

(ii) Primary education

(iii) Gymnastic for school

(iv) The teaching of mathematics

are arranged according to KWIC index as:

The : BINOMIAL theorem

Primary : EDUCATION

: GYMNASTICS for school.

The teaching of : MATHEMATICS

: PRIMARY education

Gymnastic for : SCHOOL

The : TEACHING of mathematics

The binomial : THEOREM.

In the above examples, words not required as index points are supposed to be placed on a 'stop list' and the computer is programmed to ignore them. Words such as 'the', 'of', and 'in', i.e., articles and prepositions are functional words. Remaining are content words and treated as entry elements.

A number of modified versions of KWIC have been developed. These include:

(i) Keyword out of Context (KWOC)

(ii) Keyword And Context (KWAC)

(iii) Word and Author Context (WADEX) etc.

Keyword out of Context (KWOC)

In this, keyword extracted from the title, is used as a separate heading and in the title keyword or lead word is not repeated, but is replaced by an asterik, e.g.:

BINOMIAL	The * theorem.
THEOREM	The binomial *.

Keyword and Context (KWAC)

In this sytem, lead word of the title is written at the left side of the title; e.g.,

BINOMIAL	The binomial theorem.
EDUCATION	Primary education.
GYMNASTICS	Gymnastic for School.
MATHEMATICS	The teaching of mathematics.
PRIMARY	Primary education.
SCHOOL	Gymnastics for School.
TEACHING	The teaching of mathematics.
THEOREM	The binomial theorem.

Word and Author Index (WADEX)

In this name of the authors are treated as key-words.

It is clear from above discussion that H.P. Luhn's creative ideas are greatly influenced by linguistics and especially semantics. Contextual criteria is akin to input some syntactical structure among the keywords.

3.19. Mortimer Taube

To solve the perennial problem of pre-coordinate indexing - the problem of complex concept and how to cite it, a completely different approach was developed which is known as 'Post - Coordinate Indexing System'. In this system, the indexer breaks up the complex concept into its components for indexing, the searcher then can search on any one of the components in any order. Mortimer Taube is generally considered to be the inventor of coordinate indexing (now better referred to as post-coordinate indexing, to distinguish it from pre-coordinate indexing). He was soon followed by others who developed and expanded on the concept; the whole field of computer-based indexing and searching in its present form would literally be unconceivable without it.

In the post coordinate indexing system, a new technique of term indexing entry was adapted in the place of item entry. In the term entry or single word entry system, each discrete record is for a term and not a document. Such a record bears a single term, and listed under that term are the identities of all the documents to which that term has been assigned in indexing. Such identities usually consist of simple 'accession numbers', each number identifying a single document in the collection.

One of the best known system of this kind is the 'Uniterm System' devised by Mortimer Taube. In this system:

> Each document is assigned a unique number by entering it in an 'accession register' and that numnber thereafter identifies that document for indexing purposes. A file of specially ruled cards is used, with each card bearing one of the terms from the indexing vocabulary. The ruling provides for 'terminal digit position' by dividing each card into ten columns which are numbered 0 - 9. When a document is indexed it is examined for its subject content in the usual way and suitable terms from the vocabulary are assigned to it. The card for each of those terms is extracted from the file and the number of the document is posted to all those cards. The final or 'terminal' digit in the number determines the column to which it is posted. When a search is made the searcher determines which terms best describe his needs and withdraws the cards for those terms. He then compares the numbers on the cards in order to detect which numbers are common to all cards. Terminal digit posting helps number comparison. All common numbers identify relevant documents. (Taube 1953: 57)

The following figure shows typical uniterm cards for the title 'Public Libraries in India':

Public Libraries in India

PUBLIC

0	1	2	3	4	5	6	7	8	9
10	11	22	13	14	15	6	17	8	19
30	51	72	43	24	35	36	107	38	309
40	311	92	63	304	905	106	317	98	509
50	561	502	363	634	1115	316	987	518	939
	761	812	593	1924	1225	976	1817	988	1009
1100	1001	1032	1503		1305	1006			

LIBRARIES

0	1	2	3	4	5	6	7	8	9
30	111	72	33	104	55	16	77	18	19
40	331	92	613	364	155	306	317	38	39
	551	332	553	514	335	916	937	938	99
130		852	1003	1184	605	1016	1007	1238	1009

INDIA

0	1	2	3	4	5	6	7	8	9
50	11	72	33	34	15	66	37	88	39
90	21	132	63	94	35	96	57	1008	99
100	31	852	303	134	95	116	87	1118	1009
500	1001	1002	863	1394	1215	1216	1317		

In these three cards common number is 72, therefore Acc. No. 72 document is required document.

To avoid confusion and false retrieval, Taube suggested two devices which can be utilized to reduce false retrieval. These devices are: link and role. Links show that two terms are actually treated in relation to each other in a document, rather than merely appearing together fortuitously, while roles show the special meaning of term that is used. In other words, the intent of links is to divide up the concepts into two or more parts, while roles show relation. The example, 'lead coatings for copper pipes' requires the use of roles and not that of links to prevent retrieval in search of 'copper coating' or 'lead pipes' while permitting retrieval of 'coating for pipes'. (Taube 1961: 98-100)

Analysis of coordinate indexing from linguistic point of view, it can be said to be a pioneer in introducing syntactic structure on an indexing system at the search stage. "Content needs" and "Cross-references" form part of semantics of the indexing language propounded by Taube.

3.20. Calvin N. Mooers

Calvin N. Mooers, Preesident of the Zator Company is famous for his efforts towards the study of theoretical investigations on retrieval of information in the USA.

Whereas most of Taube's theoretical investigations were directed towards supporting the concept of the aspect card development, Mooers concentrated on document organized system with his coding system. Taube was greatly concerned with the concept of coordination. Mooers adapted reverse approach to this method by listing the subject terms for a single document on a single card and, in doing so, has been concerned with the question of efficient storage of the subject information. To achieve this, he proposed the concept of storing the codes for all of the subjects related to a document in the same invariant place – one superimposed on the top of the other – the concept of "superimposed coding". The identification of each subject is thus recorded in the same field in exactly the same physical location. He first become known in 1947 for 'Zato-Coding' a superimposed method of coding by random numbers. Later on, when the use of this method was experimented in various specialized fields, Mooers defined a certain number of notions, e.g., descriptors, interlocking sets, filtering principles, and devised a mathematical theory of information retrieval. (Mooers 1959: 1346). This made him famous.

Mooers, descriptors method consist in devising for each user a 'dictionary of notions answering his

particular needs. Each descriptor represents an idea or a concept, generally fairly broad and carefully defined by a 'scope note' or a note explaining its precise use, its particular significance in the specific individual system of which it is part. The filtering technique consists in an examination, for each document, of the complete list of descriptors incorporated in the system, with a view to retaining those which may characterize the document: this technique is simplified by the use of a list which regroups the descriptors under a series of "chapter headings" presented in the form of questions.

While discussing these descriptors, he emphasizes that concept should be given priority. Mooers says that, "The descriptor's conceptual units may contain one word or several words. Descriptors are based on unit concepts not on unit words." (Mooers 1963: 24).

Analysing the ideas of Mooers, we find that his ideas are basically related with semantic part of linguistics. His definition of the notions or descriptor helps the user to realize the correct meaning of the term with which it is related.

3.21. James W. Perry, Allen Kent, and Thomas H. Rees

Perry and his colleagues at the Centre for Documentation and Communication Research, Western Reserve

University, Cleveland, Ohio, concentrated from the beginning of 1950s on semantic analysis of complex terms into individual terms and the result appeared for the first time in his book, Machine Literature Searching (Perry and Kent 1956) in the form of semantic factors. By 1958, when the book, Tools for Machine Literature Searching (Perry and Kent 1958) appeared, the scheme of factoring had been formalised. The objective of the scheme were as follows:

> The important aspect of the contents of documents be explicitly symbolized in a consistent manner. In particular, relationships must be indicated in a uniform manner, regardless of the various form of phrases used in current written language. (Perry and Kent 1958: 136-50)

Perry and his associates concentrated their contribution in the field of Semantics and Syntax. Important characteristics of semantic factoring can be enumerated as follows:

(1) The distinction between 'synthetic relationship' and 'analytic relationships', the first of these assigns special code-words of three letters and the second expresses by a variable letter (the second), within a code word of four letters indicating a semantic factor. The authors have explained these two types of relationships as follows:

> Synthetic relationship: a relation existing between concepts which pertains to empirical observation, such relationships are involved, not in defining concepts or terms, but in reporting the results of observation and experiments.
>
> Analytical relationship: the relationship which exists between concepts (and corresponding terms) by virtue of their definition and inherent scope of meaning. (Perry & Kent 1958: 76)

(2) The code words have been used to transform 'telegraphic style analysis' of documents into 'coded analyses' (encoded abstracts) containing all the essential data in the document.

(3) Special punctuation symbols are used to separate certain parts of the final coded analysis in order to avoid confusion.

(4) Hierarchical classification has been used for geographical locations and classes of objects or of concepts further subdivided by means of decimal notation.

With this objective in mind, a certain number of synthetic relationships were defined as being the most effective in recording the important aspects of the information contained in documents relating to metallurgy. These relationships were indicated by a set of symbols called role indicators.

In all twenty three role indicators were recognized in the field of metallurgy. They were grouped into five categories as follows:

(1) Relationships for materials;

(2) Relationships for properties;

(3) Relationships for the processes;

(4) Relationships for conditions;

(5) Miscellaneous role indicators.

Thomas H. Rees Jr. added nine other role indicators while experimenting abstract encoding with New York Times article (pp. 176-88). They were:

(1) Acting person or organism;

(2) Category of organism;

(3) Position (within the organism) rank or title of the person concerned;

(4) Subordinate organism;

(5) Client or user;

(6) Source of information;

(7) Location;

(8) Specified quantity; and

(9) Type of article.

These nine role indicators can be said to belong

to the category 'relationships for organisation and person'.

Analytical relationships were distinguished ten in number. They are as follows:

(A) Categorical

The word is a number of the class represented by the descriptor: e.g., Red is a colour. This is what I have called the generic or class relation.

(E) Intrinsic

The word is a whole of which the descriptor is a component; e.g., Diamond is composed of carbon. This is whole part relation.

(I) Inclusive

The word is a component of a whole represented by the descriptor; e.g., Wing is a part of bird. This is a part-whole relation, inverse to (b).

(O) Comprehensive

The word is a composite of several members of the class represented by the descriptor; e.g., a Mosaic is made up of crystals.

(U) Productive

The word produces, or is used for or to, the

descriptor; e.g., Gelation produces a celloid, a club is used to beat. This is an agent or agency relation.

(Q) Affective

The word makes use of, is determined by, or is influenced by the descriptor, e.g., Electroplanting makes use of electricity, Phytotropism is influenced by light.

(W) Instrumental

The word is produced by acts on or is acted on by the descriptor, e.g., Olive oil is produced by a plant, Hardening acts upon metal, Jacketed is acted on by cover.

(X) Negative

The word represents something which is absent in the descriptor.

(Y) Attributive

The word represents something which is represented by the descriptor, e.g., Modified is characterised by change.

(Z) Simulative

The word resembles, but is not, the descriptor, e.g., Whale and fish. (Perry & Kent 1956: 278-9)

Thus we find that semantic analysis of 'complex terms' into 'individual terms', sound approach by J.W.

Perry and his associates in generic, and coordinate relations between descriptors allows the enquirer to broaden his search in order to increase the recall factor.

3.22. Douglas John Foskett

One of the most prominent figures in contemporary British Library and Information Science, Douglas John Foskett is well known about his interest for classification. His contributions are mainly based on syntax. He was one of the founders of 'Classification Research Group' in 1952, who introduced Ranganathan's ideas in U.K. Foskett was highly influenced by the idea of Faceted Classification and tried to implement it in his special schemes of classification. In 1955, CRG passed a memorandum 'on the need for a faceted classification as the basis of all methods of information retrieval' (CRG 1955: 212-68). This led to the identification of two problems:

(i) Organization of concepts into Facets (or Facet Analysis); and

(ii) Arrangement of facets in Helpful order (Citation Order of Facets).

Again, in 1963, CRG made a proposal to develop a new general scheme of classification which could be suitable for shelf arrangement and also for use in computer based system. (Library Asso., 1964).

To meet the objectives of the CRG's memorandum and its proposal for a general scheme of classification, Foskett tried to find out solution of the problems of organization of concepts and their citation order. The search started with the proposition that there are certain kinds of ideas, such as 'things', 'properties' and 'actions' which are fundamental to a subject building system. In this direction, first attempt was made to categorize the concepts into 2 parts, e.g.:

(i) Things;

(ii) Non-things.

Things were divided into two kinds:

(i) Naturally occurring entities; and

(ii) Artificial entities.

Again, artificial entities were divided into two parts:

(i) Concrete objects or artefacts;

(ii) Abstract construct or Mentefacts, i.e., imagination, belief, etc.

Non-thing concepts were divided into two parts:

(i) Attributes of things;

(ii) Activities of things.

After categorization of concepts, attention was paid towards the problem of citation order of these categories with a view to establish a linear sequence of concepts. Various attempts had already been made but none proved to be satisfactory. The most attractive proposal in this direction was made by D.J. Foskett, who suggested that "Integrative Level Theory might provide a logical basis for linear order" (Foskett & Palmer 1961: 139).

Integrative Level Theory suggests that there is a recognizable order in nature which consits of a progression from lesser to greater levels of organization:

> This theory postulates that a sequence of entities can be established according to their different levels of organization, a new level being recognised as the point at which entities from lower levels come together, acquire a new identity, and are characterised by properties which are not found in the entities at lower levels. Molecules, for example, possess properties and structural characteristic which are not found at the level of their constituent atoms, while unicellular living organisms possess attributes over and above those of their molecular components. (Austin 1976: 174).

Based on this theory, Foskett and his colleagues identified the sequence of entities as follows:

Physical Entities

Levels I Fundamental particles

II Atoms, isotopes

III Molecules

IV Molecular assemblage, i.e., solid.

Chemical entities

Levels I Elements

II Compounds

III Complex compounds

Heterogeneous Non-living entities

Levels I Raw material

II Processed raw material

III Components

IV Finished articles.

Artefacts

Levels I Minerals

II Rocks

III Physiographic features

IV Astronomical entities.

Biological entities

Levels I Viruses

II Organelles

III Cells

IV Tissues

V Organs

VI Systems, i.e., digestive system

VII Organisms

VIII Communities, e.g., shoal, herd.

Man

Levels I Individual

II Group

III Local community

IV National community

V International community

Mentefacts

Levels I Units, e.g., digit, note

II Words, numbers, bars, etc.

III Sentences, formulae, musical phrases etc.

IV Paragraphs, themes, etc.

V Complete works, philosophical systems etc.

(Foskett 1970: 30-31)

Later on, when this theory was applied by Foskett and his colleagues to properties of entities and their

activies, it was soon realized that this theory could not be applicable. Therefore, another principle of Decreasing Generality or Increasing Specificity was adopted and the result in case of the purposes was as follows:

1. General relative terms

    Degrees of amount

    Degrees of order or rank

    Degrees of kind or substance

2. General evaluative terms

    Favourable reaction terms

    Unfavourable reaction terms

3. Positions

    Time

    Space

    Person as user or possessors

4. Physical Measure

    Mass

    Linear measure

5. Shape

6. Appearances

    Light

    Colour

7. Sounds

8. Tactile sensations

9. Flavours

10. Odours

11. States

Mechanical states

Energy states

12. Structure (Foskett 1970: 34-35)

In the same way, the application of the Principle of General to Specific to activities made static conditions preceded by the dynamic and the result was as follows:

1.1. General Activity concepts.

2.2. General static and kinetic conditions

3.3. Equilibrium

4. General kinetic conditions

5. Contacts and disturbances

6. Motions and transfers

7. Assembly and disassembly. (Foskett 1970:36)

On the sequence of concepts in subject statement, Foskett suggested that 'General System Theory' may provide a framework within which the component parts of any subject analysis could be ordered consistently. Later on,

this idea found on concrete shape in the following role operators as suggested by Derek Austin:

(1) Properties of system
(2) Second system/Environment related to (3)
(3) Activities and interactions not involving material change.
(4) Active sub-system
(5) Passive sub-system
(6) Interaction with system
(7) Second system/Environment related to (8)
(8) Interactions involving change of system
(9) Attribute defining a sub-system.

According to these operators, the stages in structuring compound subjects become:

(a) distinguish and set down the various concepts discerned, designating them as entities, properties and activities.
(b) designate entities as passive system, active system(s) or sub-system(s). Class at the passive system, indicating others with appropriate operators.
(c) assign activities to the entities to which they apply, and properties to either entities or activities.

(d) consider the attributes which in principal focus of the documeent, and indicate that by the appropriate operator.

(e) indicate all other attributes by adding(s) to the operator which declares which entity possesses the attribute or which activity possesses the property of an activity. (Foskett 1970: 41)

In addition to these theoretical contributions, Foskett designed many special classification schemes. Out of these schemes, Container Manufacture, Food Technology and Health and Occupational Safety are the most important. The categories used in these schemes are as follows:

Names of Classification Scheme	Categories (6)
1. Metal Box Container Manufacture Company's Classification System	(A) Four for manufacture of Boxes: (i) Products; (ii) Parts; (iii)) Materials; and (iv) Operations. (B) Two categories for packing and crating:

Names of Classification Scheme	Categories (6)
	(i) Packed and crated products; and
	(ii) Material condition of the latter; processes.
2. Food Technology Classification Scheme	(i) Products;
	(ii) Parts;
	(iii) Materials; and
	(iv) Operations.
	(Vickery 1975: 156-61)
3. Health and Occupational Safety Classification Scheme	B. Physical agents and natural phenomena.
	C-G Substances.
	H-J Premises, equipment, processes and operations.
	K Organisation of labour and industrial structure.
	L Fire and explosions.
	M-N Pathology.
	P Physiology and Psychology.
	Q Research techniques.
	R Medical prevention and treatment.
	S Techniques of safety and health.

Names of Classification Scheme	Categories (6)
	T Equipment for individual protection.
	V Organisation of Safety and health.
	W Categories of persons.
	X Industries
	Y Special aspects
	Z Generalia.
	(Foskett 1960: 102-7)

Thus we can say that Fosckett implemented the ideas of Ranganathan's Faceted approach as well as his own which are based on syntactic level of linguistics.

3.2.3. Barbara Kyle

Classification Research Group (CRG) of U.K. was highly impressed by the principles of facet analysis and synthesis of Ranganathan. During 1950s, CRG decided to implement these principles in the preparation of special schemes of classification. Barbara Kyle's attempt to prepare a special classification scheme for social science is one of these attempts.

The main part of Kyle's classification consists of the two facets, Activities and Personalities, denoted

respectively by upper and lower case letters. In classifying a document, the Activities facet comes first, contrary to the sequence in CC and in the schemes made by other members of the CRG, but in the line of social scientists.

The personalities range from individuals, subdivided according to age, sex, relationships and so on up to states and international organization. She has applied the 'principle of integrative level' for the arrangement of entities in the facet of personalities. According to this principle, there is a recognizable order in nature which consists of a progression from lesser to greater level of organization, e. g., for arrangement in social sciences we can arrange these in the following way:

Man

Level	I	Individual
	II	Group
	III	Local community
	IV	National community
	V	International community.

The 'principle of general to special' has been followed for the arrangement of entities in the facet – Activities. The following illustrates it:

Activities from Education

Ha	Education, General
Hab	School organization
Ham	Teaching method
Hat	Curriculum
He	Types of education
Hed	Vocational education
Heg	Technical education
Hem	Physical education etc.

While discussing relationships between the general order of her classification she emphasized that the arrangement of the main classes must vary according to the user. She advocates, "The order of the main classes in my table is that prescribed by the interest of users" (Kyle 1958: 183).

Observation reveals that Kyle's Classification produces a better sequence of entries, with more related topics falling close to each other and offers a far superior indexing technique for retrieval. Structurally, its significance lies in the extension of the technique of facet analysis into the literature of social science and the interesting modifications Kyle has made to the generalized facet formula given by Ranganathan. Syntax

part of linguistics can be discernable in the application of facet formula and citation order whereas semantic part can be seen in the 'principles of integrated theory' and 'principles of general to special'.

3.24. Jean-Claude Gardin

The subject indexing so for considered have begun by selecting Main Classes and proceeded to divide them, from the general to special. An alternative method to the above approach was adopted by J.C. Gardin, who studied the analysis of relations between concepts, building up classification schemes from the individual concept to the general class. Gardin's theory of relational analysis is based on modern work in structural linguistics.

The contribution of Gardin is in the field of "establishing a general system of logical reduction of language, suitable for specialized literature in the humanities" (GEDSH 1960: 1). This research work was entrusted to Gardin by 'EURATOM'. The findings of this work were reported in the reports of the Groupe de Etudes Documentaires pour les Sciences Humaines of the Association March Bloch. The result of this work was Syntagmatic Organization Language, or SYNTOL.

Gardin's original work was in Archaeology where he devised a scheme for the classification of artefacts,

and quickly realised that it would work with a simple punched card system for the coding of documents. This work has now been extended to a general theory, stemming to a large extent from modern work in structural linguistics. There are two main forms of organization in SYNTOL: the 'Paradigmatic' and the 'Syntagmatic'. Paradigmatic consists of a basic, 'fixed' relational linkage between two terms, as for exmaple, in a simple hierarchy. The Syntagmatic expresses the relation between terms as found in the literature.

The most significant feature of 'SYNTOL' is its explicit differentiation between a priori relationships, i.e., those that are known in advance and a posteriori, i.e., those which are found only by scanning a particular document.

Gardin used the term semantic to denote a priori relationships, i.e., those depending on definitions of terms, and syntactic to denote the a posteriori relationships arising out of the association of terms in particular documents. Gardin coined the terms 'paradigmatic' and 'syntagmatic' to denote these two kinds of relationships, but he emphasized that the distinctions, he has in mind, is an operational one; the relationships between

two concepts may be either, depending on how we build it up into the system.

In SYNTOL we have a basic unit consisting of two terms plus a relationship 'A dyadic string in the form (R, a, b) is called a 'Syntagma'. In one form, there are four such relationships. These are:

(1) **Predicative**

Attaching to a term indicating an element or entity practically autonomous as an object of study, an essentially dependent property which describes its state, quality or function, i.e., predicate.

(2) **Consecutive**

Joining two elements of which the presence or action of one affects the presence, state or status of the other. This relationship can assume various meanings, according to the nature of the two elements concerned: the opposition of active and passive, a genetic relation, causal relation, conditional relation, functional or factorial relation, bond of finality, mediatory relation.

(3). **Associative**

Joining two elements of which the mention of one implies the simultaneous mention of the other. This

relationship also assumes different meanings according to the context: relation of the part to the whole, of place, of hierarchy, of appurtenance, of specification, semiological relation.

(4) Comparative

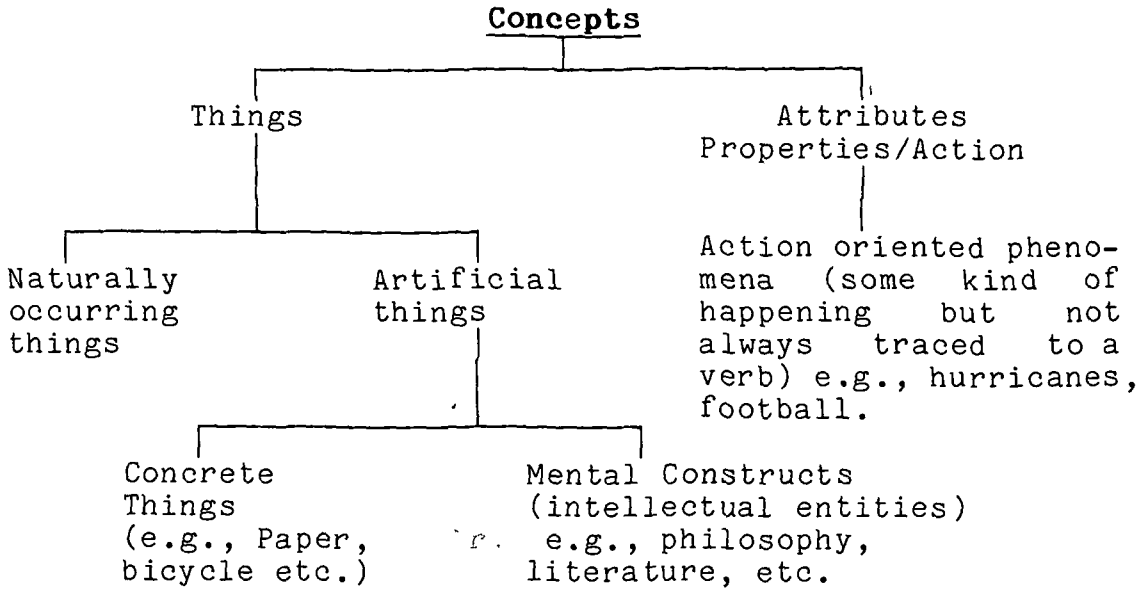
Indicating an extrinsic bond between two elements. (Grolier 1962: 88).

3.25. Derek William Austin

Austin made significant contribution to the field of indexing through his work with the Classification Research Group (CRG) and the development of PREserved Context Index System (PRECIS) (Austin 1984), a name that has become synonymous for Derek Austin. In 1976, he received the first 'Ranganathan Award' presented by FID for original contribution to classification. The contributions of Austin prominently lie in the dimension of Syntax.

For the organization of concepts, CRG, in 1960s, conducted research into the theoretical basis for a new freely faceted classification in which they tried to organize concepts into basic classes without regards for the discipline with which they are usually associated. For this purpose, two basic classes of concepts were

proposed: Things and their attributes. These two classes were further divided as follows:



Similar ideas have been proposed by other library scientists too, e.g., the citation order in Kaiser's 'Systematic Indexing' depends upon a prior distinction between 'concrete' (= Things) and 'processes' (= actions). The same division has been recognized by Coates; it occurs again in PRECIS; it is also offered as a basis for testing the validity of hierarchical relationships in the British Standard on thesaurus construction. The same distinction is encountered again in linguistics, even though the names are different. Chafe for example, based his search for linguistic universals upon the prior assumption that:

> The total human conceptual universe is dichotomised initially into two major areas. One, the area of verb embraces states (conditions, qualities) and events; the other, the area of the noun, embraces 'thing' (both physical objects and reified abstractions. (Austin 1982: 87-88)

It should be noted here that Chafe considers conditions and qualities (these generally correspond to the 'properties' in indexing languages) as closely related to verbs (= actions), not as a separate class.

Again he agrees with the ideas suggested by some documentalists that:

> Natural languages (with particular reference to linguistic universals) should be seen as the source of the basic relations and general categories that underlie indexing languages. This hypothesis is attractive, insofar as it offers a plausable explanation of the many observed correspondences between the grammatical categories etc. studied by linguists and the roles, relations and concept classes employed in classification and indexing languages. This certainly applies to PRECIS. Some current studies indicate that the principal role operators used in this indexing system can all be explained in terms of deep cases and the special classes of action organized by linguists. (Austin 1982: 93)

Austin summarises the concept 'classes' and their relationships as follows:

> (i) The principal operators that indicate the syntactical roles of terms in PRECIS can be explicated through reference to linguistic categories such as deep cases ... The same kind of explication, with similar result, has been applied to other documentation languages.
>
> (ii) Relations can be detected not only in the syntagmatic relations studies by linguists, but also in the paradigmatic relationships that are generally regarded as falling within fields such as philosophy and theoretical classification. (Austin 1982: 93-94)

He implemented these ideas in PRECIS. Researches of 1970s even pushed Austin to the point of 'translingual' indexing i.e., the ability of PRECIS to analyse and represent semantically consistent index statements in any language by looking through immediate grammatical construction to deep case grammar - the 'real' relationship of concepts at a level expressed below the natural language.

PRECIS was adopted by British National Bibliography for producing entries in the subject index from January 1971 to December 1973.

The system in BNB functioned as the product of two different but interdependent teams, each of which produced input for computer held files. The first team concerned with organizing the terms into strings based

on their syntactic relationships which led to the production of sets of index entries, while the second engaged in building up hierarchies of references based on semantic relationships between individual concepts, which ultimately led to the production of machine - held general thesaurus. Between them, these two systems constituted all the verbal sides of the index system — one of them expressed a summary statement of the subject at any level of specificity, and the other guided the user to an appropriate entry word from other terms which shared an element of common meaning. (Austin 1972: 111). In other words, we can say that the PRECIS system has two aspects, the syntactic and the semantic. In syntactic aspect, terms are organized into strings using role operators which are then manipulated by the computer in such a way as to produce entries, each carrying with it the context of the item being described. In semantic aspect, there is firm control of terminology in the index, through the use of a well-organized thesaurus.

3.26. Eugene Garfield

A variety of products, ideas, techniques and Institution (ISI - Institute for Scientific Information, Philadelphia, Pennas) can be identified with Garfield, but just two products, 'Current Contents' and 'Science

Citation Index' and their conceptualisation and development, their implementation and successful management, would stand alone as major monuments to his skills, energy and intellect. 'Current Contents' capitalizes upon the researcher's urgent needs for current research information and the human's enormous capability to scan text skillfully. 'Science Citation Index' identifies citation behaviour as a means of acknowledging intellectual debt and assigning credit, as the key to ascertain basic organizational features of the scientific literatures, useful for retrieval, for research on the nature of science, and for science policy.

The citation indexing method was developed by Eugene Garfield in the 1950s. He recognized that there was an important author-designated link between a document and each bibliographic reference cited in it. An author's references to earlier works reveal that these are documents relevant to the current one. These in-built links are exploited in citation indexes. (Garfield 1955: 103 -11). This shows the priori relationship which expresses semantic relationship. His main contribution belongs to the field of Semantics.

The citation index has two parts, the source

index and the citation index. The source index lists all documents included in the journals covered by the index. The citation index lists authors' name in alphabetical order, the citations made in any of the source documents, and indicates for each cited document all source documents which cite it (i.e., all documents which cite the documents are listed). The best known citation indexes are Science Citation Index, Art and Humanities Citation Index and Social Sciences Citation Index.

The basic idea of citation indexing is to bring all the documents which have included a given item in their list of references. Books and articles normally provide citations which are referred to. In other words, there is an inverse relationship between the original document and the document citing it.

Citation index is very helpful tool of retrieval. One can choose the relevant articles of one's interest, and others which are not relevant, can be discarded. It helps in measuring the impact of current research and spots the field where the rate of progress is high. Citation index can function as a science management tool. It can reflect the activities of science - its structure and development.

The advantage of citation indexing is that no human indexing is required and there is no need to find words to express concept. On the other hand, the retrieval performance of the citation index depends on the accuracy and consistency of the citations which cannot entirely be relied upon.

3.27. Eric J. Coates

To find out solution and criterion for the problems of entry element citation order in compound subject heading, different authors have suggested different criteria and approaches. Important among these are: Cutter's concept of 'Significance', Kaiser's formula of 'concrete/process', Ranganathan's 'set of categories' and Farradane's categories of relationships etc. If we accept these above stated approaches, one has to decide which component is more significant and therefore, should be accepted as entry element in the compound subject heading.

To solve these problems Coates in 1960, published the most significant book on cataloguing, Subject Catalogue: Headings and Structure. Some years later he put some of his ideas into practice when he became the first editors of British Technology Index (BTI). BTI headings are alphabetical but based on classification principles.

The 'things' and end product (the personality facet in Ranganathan's term) is always the key term in the entry followed by terms denoting part, property, material and action. Cross references are made from all other terms which are considered to be sought headings.

The main contributions of Coates are: to find out significant terms in a compound heading and citation order of these components. Both these problems are linguistic oriented. Significant components can be divided by means of meaning whereas the citation order depends on relationships between these terms.

On the significant terms in the compound headings, he says:

> A word which evokes a static image is more significant than one which denotes actions or processes. A static image is produced by names of 'things' and names of 'materials'. The difference between the two from the point of view of the imagination is that a 'thing' has a boundary, while a 'material' has none. For this reason we must rate the name of the material lower in significance than the name of a 'thing'. On the other hand, the image of a material is made up of entirely static — seeming properties, such as colour, hardness, smoothness, so we must rate it higher in significance than the name of action. Thus the following order of significance has been set up: Thing/Material/Action. (Coates 1960: 51).

This significance order is influenced by the Kaiser's order of concrete/ processes. Coates expanded the concrete of Kaiser into three parts, namely, things, part and material. Thus he prescribed his significance order as:

Things/Part/Material/Action.

To decide the citation order of components in a compound subject heading where each component is equally significant, he proposed the 'term of relationships' as the determining criterion. The use of the analysis of relationships between components as a means of determining component order owes its origin to Farradane. To show relationship among terms, he has devised twenty different kinds of relationship among terms. "The simplest of all relationships between concepts is represented by the genitive 'of'. For subject heading purposes, we reverse the order of the concept from that in which they stand in the amplified prepositional phrase" (Coates 1960: 52), for example:

(i) Exploitation of invertebrates becomes in subject heading as:

INVERTEBRATE, Exploitation (Thing/Action)

(ii) Deterioration of oil transformers becomes

TRANSFORMERS, Oil, Deterioration (Thing/Material/ Action).

There is an exception to this rule where 'of' is used for 'made up of' and connects name of a thing to the name of a material. In this case, the order of the subject heading will remain same as prepositional phrase in natural language, e.g.,

'Table of wood' will be TABLE, WOOD.

To use 'for' to show relation between concepts, the same rule should be followed as that of preposition 'of', e.g.,

(i) Container for food = FOOD, Container

(ii) Telescope for Astronomy = ASTRONOMY, Telescope.

There are some exceptions to the above rule where the order of the natural language will remain same. It will not change in the subject heading, e.g.,

(1) When 'For' is used to distinguish homonyms.

(a) Counter for (a coloured disc), game = COUNTER, games.

(b). Counter for business equipments = COUNTER, Business Equipments.

(2) If 'for' is used for pedagogical literary relationship, e.g.,

(a) Physics for Engineer = PHYSICS, Engineer

(b) Social Science for Nurses

= SOCIAL SCIENCES, Nurses.

(3) When 'For' is used to express method of operations and not as purpose, e.g.,

Library for Lending = LIBRARY, Lending.

Phrases where two nouns stand in the same heading, the citation order will be as in natural language.

If they belong to noun and adjective, adjective should be changed into noun and table rule will be followed, e.g.,

Comparative anatomy

= Comparison anatomy (I stage)

= Comparison of anatomies

= ANOTOMY, Comparative.

Apart from these, Coates also prescribes citation order for subject and locality. He suggests to solve this problem by ranking of the main areas of knowledge to the extent to which these appear to be "significantly conditioned by locality". On the question of place and

subject he has devised the following ranking order:

(i) Geographical and biological phenomena (Flora & Fauna).

(ii) History and Social phenomena.

(iii) Language and Literature.

(iv) Fine Arts.

(v) Philosophy and religion.

(vi) Technology.

(vii) Phenomena of physical sciences.

He has suggested that serial Nos 1-3 should be entered under place, except the individual species in botany, and works that are studies of these subjects, e.g.:

'Teaching of Sociology in Australia'

will go under subject. Subject place combination have been suggested for such subjects.

Coates' approach to solution of 'significant elements' and 'citation order' are largely dependent on linguistics. His 20 kinds of different relationship are clear examples of this statement. In many cases he has proposed that the citation order of natural language should be maintained as it is. Even in the case of signi-

ficant term formula, extension of the grammatical categories of natural language: Subject/predicate or things/ material action has been suggested.

3.28. B.C. Vickery

Vickery, as one of the important members of Classification Research Group (CRG), took keen interest in the field of Information Retrieval, specially in classification. He was highly influenced by the idea of Ranganathan's Faceted Classification and tried to simplify it in the following way:

(1) A whole collection (universe) of entities may be divided into a series of fields (For example, Physics into Mechanics, Optics, Heat, Electricity etc). Each field is, therefore, broad grouping. The fields may be arranged in some meaningful sequence.

(2) Each field may be divided into a series of facets (for example, Chemistry into Substance, Properties, Reactions etc). Each facet is therefore grouping. The facet in each field may be arranged in some meaningful sequence.

(3) Each facet may be structured into a hierarchy, subdivided stage by stage by the application of a series

of characteristics. These are applied in ordered sequence. Any sequence of term generated by successive subdivisions is a chain (for example, the chain 'Machines - Mechanical and thermal machines - Nuclear reactors - Heterogeneous reactors - Thermal heterogeneous reactors - Graphite moderated - Gas coded'). Any one level of subdivision gives rise to a group of terms that constitute an 'array' (for example, within the class. Metal array ('Bismuth, Lithium, Mercury, Potassium, Sodium'). The term in each array may be arranged in some meaningful sequence.

(4) Rules may be provided for combining (co-ordinating) terms from the same array, from different arrays in the same facet, from different facets into the same field, and from different fields. These rules may involve the use of special relational operators or role indicators.

(5) Each field, facet and term may be coded to fix its position in the whole system and to facilitate unambiguous combination with other codes.

(6) An alphabetical index to the terms, facets and fields may be provided, showing the code of each. (Vickery 1986: 150).

However, he did not agree with Ranganathan that there could be only 'Five Fundamental Categories'. He is of the view that categories may differ from discipline to discipline. He designed many special classification schemes. Out of these special schemes 'Soil and Earth Sciences', scheme for Astronomy and scheme for Reactors are important. He formulated eight categories for the scheme 'Soil and Earth Sciences', nine for the scheme for Astronomy and ten categories for the scheme 'Reactors'. In spite of this he has provided a list of 9 categories which may be applicable in general to all disciplines. These 9 categories are:

1. P Substance, Product, Organism.
2. O Part, Organ, Structure.
3. C Constituent.
4. Q Property and Measure.
5. R Object of Action, Raw Material.
6. E Action, Operation, Process, Behaviour.
7. A Agent, Tool.
8. G General Property, Process, Operation.
9. ST Space and Time.

(Vickery 1970:)

Further, to arrange the terms in helpful sequence in a subject statement, Vickery has suggested 'Standard

Citation Order". This provides the most important relationship among terms with particular reference to scientific literature. The standard citation order postulated by him is as follows:

> Thing - Part - Constituent - Property - Measure - Patient - Process/Action/ Operation - Agent - Space and Time. (Vickery 1958: 35)

This order of terms is vital from the point of view of syntactic relation. His categorization of terms into facets, may be regarded as semantic part of his indexing language.

3.29. Jean Aitchison

The thesauri of 50s and early 60s were structured purely alphabetically. The limitations of the alphabetical arrangement led to the employment of classification aids, ranging from the broad to the detailed, and from auxiliary to integrated devices. The traditional thesaurus method of indicating broader and narrower term hierarchies has proved unsatisfactory because not all levels of hierarchy can be displayed alphabetically at one place, or even if such an effort is made, it is not possible to distinguish between different hierarchical levels. To overcome this disadvantage, a new device was designed by

Jean Aitchison, known as "Thesaurofacet" (Aitchison 1970: 187) in 1969, which integrates a detailed faceted classification with an alphabetical thesaurus. The Thesaurofacet, a thesaurus and faceted classification for Engineering and Related subjects, developed from an existing faceted classification, was first published as a Thesaurus/Classification from the English Electronic Company in 1969.

The structure of Thesaurofacet as described by Aitchison is as follows:

> The Thesaurofacet consists of two sections: the classification schedules and the alphabetical thesaurus. The terms appear twice, once in the thesaurus and once in the schedules, the link between the two locations being the notation or class number. At each location, different information is given about the term. The two parts of the system are complementary; if consulted apart, they are incomplete. (Aitchison 1972: 82)

For example, if we want to search out the term "Television Camera tubes" then we have to go via number MCE to the schedule "Electron tubes". Here, we find that the term Television Camera Tube belongs to the hierarchy of Electron Tubes. The immediate genus of Television Camera Tube is Cathode ray tubes (BT) and its genus is Electron beam deflection tubes and Electron tubes, while

the narrower term is "Television colour camera tube". There are also related terms which are species of the common genus "Cathode ray tubes". These include "Image converter tubes" and Television picture tubes.

Thus we find that classification entry can display these relations more effectively than conventional thesaurus entries.

After consulting classification schedule, if we return to thesaurus part for additional information about the terms, we find that "Television camera tubes" in the thesaurus entry shows that this term is also species of the term "Television apparatus" and that it has other related terms which might be searched as "Photo tubes" or Television Cameras which are not available in classification schedule. Thus we find that both the parts i.e., Thesaurus and classification parts are complementary to each other and are helpful in getting required information. The following complementary chart prove the above statement.

Complementary Parts of the Thesauro-facet

(I) Thesaurus Entry

Information not in Classification Schedulees

Television Camera Tube MCE
Class No.

UF Camera tubes (Television)
Emitrons
Iconoscopes
Image orthicons
Orthocons
Pick up tubes (Television
Vidicons

RT Photo multipliers
Photo tubes
Television cameras

BT(A) Television apparatus

(II) Classification Schedules

Information not in Thesaurus

Related terms (RT):

i) Image converter tubes
ii) Image intensifiers
iii) Storage tubes
iv) Television picture tubes
v) Television colour picture tubes

Narrower Term (NT):

Television colour camera tubes

Broader Terms (BT):

i) Cathode ray tubes
ii) Electron beam deflection tubes
iii) Electron tubes

M	ELECTRONIC ENGINEERING
MA	ELECTRON TUBES (Contd) BT
MBT	ELECTRON TUBES beam deflection tubes BT
MBV	Indicator tubes (tuning)
MBW	Trochotrones
MC	Cathode Ray tubes BT
MC2RT	Image converter tubes
MC4	Image intensifiers
MC6	Storage tubes
MCE	Television camera tubes
MCI	Television colour camera tubes NT
MCL RT	Television picture tubes
MCO	Television colour piture tubes
MCO	X ray tubes.

(Aitchison 1972: 83)

Analysis of his contribution reveals that semantic as well as syntactic of an indexing language fully taken care of by utilizing classification and thesaurus techniques in her thesauro-facet.

3.30. Ganesh Bhattacharyya

Ganesh Bhattacharyya is well known for his contribution to subject indexing. He devised a indexing technique called 'Postulate Based Permuted Subject Indexing' (POPSI). POPSI is based on the principles of classification according to the Ranganathan School of Thought. This helps to give context to terms in the preparation of subject strings. POPSI uses Analytico-Synthetic method for string formulation and permulation of the constituent terms. This is guided by scientific and well recognised principles and postulates.

Bhattacharyya is of the opinion that structure of a subject statement lies in three dimensions. These are:

(a) Semantic Structure

It includes Genus-species, whole part, and inter-facet relationships.

(b) Elementary Structure

It means the dimension of elementary categories to which the constituent concepts of a subject statement belong.

(c) Syntactic Structure

It stands for the sequence of constituent concept in a subject statement. (Bhattacharyya 1981: 23).

To achieve the deep structure of the subject indexing language, he surveyed many subject indexing languages and came to the conclusion that elementary categories are the only base for this deep structure. Therefore, he formulated four elementary categories, viz.:

(i) Discipline;
(ii) Entity;
(iii) Property, and
(iv) Action.

(Acronym of these Categories is 'DEPA')

Apart from these categories he also formulated a special component called 'Modifier'. The cannotation and denotation of these categories are as follows:

Discipline

Discipline is an elementary category that includes traditional fields of study.

For example: Chemistry, Agriculture, Sociology etc.

Entity

It includes manifestations having perceptual correlates or only conceptual existence as contrasted with their properties and actions performed by them or on them:

For example: Building, Aeroplanes, Stomachs etc.

Properties

This includes attributes, quantitative or qualitative.

For example: Gravity, volumes, strength etc.

Action

Action includes concept of doing. It may be self action or external action.

For example: Migration is self action, whereas selection, evaluation are external actions.

Modifier

In relation to the manifestation to any one of the elementary categories, modifier refers to qualify the manifestation without disturbing its conceptual wholeness.

For example: 'Subject' in Subject Classification.

He has also provided citation order to these elementary categories, which constitutes the syntactic structure. This is as follows:

> Discipline followed by Entity followed by Property and/or Action. Property and/or Action may further be followed by Property and/or Action as the case may be. Each of the above components may further admit of, and be followed immediately by their respective species/types and/or Parts and/or Special Modifiers. The common Modifiers generally occur last in the sequence. These rules of Syntax are in total conformity with Ranganathan's Theory of Classification (especially with the Principles of facet sequence) based on the principle of decreasing concreteness (Wall-Picture Principle) and its derivatives. (Aptagiri et al. 1995: 163-4)

His contribution in the field of semantics lies in grouping terms of subject prepositions into two types of classification: Organizing Classification and Associative Classification.

Organizing Classification groups terms/classes on the basis of four kinds of relationship. These are:

(i) Coordinate;

(ii) Super ordinate;

(iii) Subordinate; and

(iv) Collateral.

The result of Organizing Classification is always a hierarchy.

Associative classification, first identifies a group and then brings its members together on the basis of their association to a common factor.

Analysis of his contribution reveals that semantic and syntax of an indexing language have been well cared by Bhattacharyya on the basis of certain Fundamental Principlees.

3.2. Generalization

A search for principles for information retrieval based on linguistic analysis led us to the works of certain scholars, whose contributions have been universally recognised in the field. They are:

Charles Ammi Cutter, James Duff Brown, Julius Otto Kaiser, Henry Evelyn Bliss, Shiyali Ramamrita Ranganathan, Gerard Cordonnier, Douglas John Foskett, Jasua Edward Louis Farradane, Barbara Kyle, Brian C. Vickery, Jean Aitchison, Eric J. Coates, Derek Austin, Jean-Claude Gardin, James W. Perry, Allen Kent, Thomas H. Rees, Calvin N. Mooers, Martimer Taube, Peter Hans Luhn, Eugene Garfield and Ganesh Bhattacharyya.

Principles formulated by these scholars at semantic and syntactic levels are provided in Table 3.1.

A close study of all these principlees reveals that these principles can be contained in the canons, postulates and principless formulated by Ranganathan.

Ranganathan's model is a comprehensive one consisting of both semantic and syntactic principles whereas the principles developed by other scholars are either covering semantic level or syntactic level. There are very few who have tried to cover both the levels.

Most of the scholars have talked in terms of syntactic devices, whereas Ranganathan speaks in terms of certain Fundamental Categories. These are PMEST. He is of the opinion that each concept is a manifestation of one of these Five Fundamental Categories and their Rounds and Levels. He has suggested that if a subject is having more than five concepts, they can be analysed in the context of Rounds and Levels of Fundamental Categories. All these ideas of Ranganathan are seminal as contained in his Postulates and Principles.

All these categories suggested by others are reducible to Ranganathan's Five Fundamental Principles. A comparative chart is given in Table 3.2.

TABLE 3.1

Principles of Information Retrieval propounded by different scholars on the basis of Semantics and Syntax

Sl.No.	Name of Scholar	Semantic Level	Syntactic Level
1.	C.A. Cutter	1. Principles of Syndetic entries (Rules 168-171, 173, 187-188). 2. Principle of Specific entry rules (161)	Principles for usage of entries. (Rules 162-167, 174-175)
2.	J.D. Brown	'One Place Theory'	Citation order of Subject Genus-Species in Science Subject.
3.	J.O. Kaiser	Syndetic Structure Higher Collectives Lower Specifics Synonyms	1. Categorization of Subject Statement Concrete and Process. 2. Citation order (a) Concrete-Country-Process (b) Country-Concrete-Process.
4.	H.E. Bliss	1. Principles of Consensus 2. Principle of Subordination of special to general. 3. Principles of alternative location.	Citation order in classes Pure-Applied
5.	S.R. Ranganathan	1. Canons of Characteristics 2. Canons for Array 3. Canons for Chain 4. Canons for Filiatory Sequence. 5. Principles of Isolate Sequence 6. Canons for Verbal Plane.	1. Postulates for classification. 2. Principles for Facet Sequence.

S.No.	Name of Scholar	Semantic Level	Syntactic Level
6.	Gerard Cordonnier		9 Categories and their citation order i) Organisms & Services; ii) Persons (Miscellaneous categories); iii) Individuals (living beings); iv) Bodies (natural, Simple, Compound miscellaneeous conditions); v) Miscellaneous equipments; vi) Miscellaneous actions; vii) Intellectual concepts; viii) Documentary forms; and ix) Time.
7.	D.J. Foskett		Provided different sets of categories for different subjects. 1. Metal Box Co. (5) Classification Scheme. (a) Products, (b) Parts; (c) Material; (d) Operations; (e) Miscellaneous common subdivisions. 2. Food Technology (4) Categories (a) Products; (b) Parts; (c) Materials; (c) Operation.

S.No.	Name of Scholar	Semantic Level	Syntactic Level
			3. Health & Occupational Safety (16) Categories i) Physical Agents natural phenomena; ii) Substances; iii) Premises, equipment, processes & operations; iv) Organization of labour and industrial structure; v) Fire and explosives; vi) Pathology; vii) Physiology & Psychology; viii) Research Techniques; ix) Medical prevention & Treatment; x) Techniques of Safety & Health; xi) Equipment for individual protection; xii) Organization of safety & health; xiii) Categories of persons; xiv) Industries; xv) Special aspects; xvi) Generalia;
8.	J.E.L. Ferradane	Abandoned the idea of main or traditional class	Role Operators (9) 1. Concurrent 2. Equivalent 3. Distinctness 4. Comparison & Self Activity

S.No.	Name of Scholar	Semantic Level	Syntactic Level
			5. Dimensional stage 6. Reaction 7. Association 8. Appurtenance 9. Causation or functional dependence. <u>Four categories for subject classification</u> (a) Entities (b) Activities (c) Abstracts, & (d) Properties.
9.	Barbara Kyle	Abandoned the idea of main class or traditional class	(1) Two categories of social science, Personality and activities. (2) Citation order for these categories as: Activities - Personality.
10.	B.C. Vickery		Provided different sets of categories for different class and provides standard citation order of these categories. However, he provides 9 categories for all purposes: 1) Substance, product, organism. 2) Part, organ, structure. 3) Constituent. 4) Property and measure. 5) Object of action. 6) Action operation process behaviour.

S.No.	Name of Scholar	Semantic Level	Syntactic Level
			7) Agent, tool. 8) General category, process, operation 9) Space and Time. The above sequence is the order of their sequence as they are listed.
11.	Jean Aitchison	Equivalence Relation Hierarchical Relation Associative Relation One place per concept philosophy.	Facet analysis - 9 categories with their citation order 1) Industries 2) Machines and Systems 3) Ancillary plant & components 4) Materials 5) Physical phenomena 6) Operations 7) Agent (Instrumental & Equipments) 8) Languae & form division 9) Geographical division.
12.	E.J. Coates	Cross references	4 Categories Things/Part/Material/Action. This shows their citation order also.
13.	Derek Austin	Thesaural relationship Paradigmatic relationship based on De Saussure's Paradigmatic relation.	1. Role operator of PRECIS. 2. Syntagmatic Relation based on De Saussure's Syntagmatic relationship.
14.	J.C. Gardin	Paradigmatic relation.	Syntagmatic relationship.

S.No.	Name of Scholar	Semantic Level	Syntactic Level
15.	J.W. Perry Allen Kent T.H. Rees	Semantic factoring	5 Categories and their citation order. (i) Relationships for materials; (ii) Relationships for Properties; (iii) Relationships for Processes; (iv) Relationships for conditions; (v) Miscellaneous role indicators.
16.	C.N. Mooers	List of descriptors Coordination (Post)	
17.	Martimer Taube	Uniterm coordinate Indexing (Post)	
18.	Peter Hans Luhn	Key-word-indexing KWIC and its various forms.	Contextual idea of Linguistics.
19.	Eugene Garfield	Associative Relationship	
20.	G. Bhattacharyya	Thesaural relationship	4 Categories including their citation order as follows: 1. Discipline; 2. Entity; 3. Property, and 4. Action.

TABLE 3.2

Schematic chart of categories by different scholars: A comparative statement.

S. No.	Name	Categories						
		Personality	Matter			Energy	Space	Time
			Material	Properties	Method			
1.	Ranganathan	Personality		Matter		Energy	Space	Time
2.	J.O. Kaiser	Concrete			Processes Services			
3.	Gerard Cordonnier	Organisms Persons Individuals Bodies Intellectual Concepts		Equipments	Services Action			Time
4.	D.J. Foskett	Products Parts	Material			Operations		
5.	J.E.L. Ferradane	Entities Abstract		Properties		Activities		
6.	Barbara Kyle	Personality				Activities		
7.	B.C. Vickery	Part Constituent Object of Action	Substance	Properties		Action/ Operation Process		

S. No.	Name	Categories						
		Personality	Matter			Energy	Space	Time
			Material	Properties	Method			
					Agent/ Tool	Gen. Cat. Process/ Operation	Space	Time
8.	Jean Aitchison	Industries Machines Components Physical phenomena	Material		Agent/ Tool	Operation	Geographical Division	
9.	E.J. Coates	Things Part	Material			Action		
10.	G. Bhattacharyya	Discipline Entity		Property		Action		

Further, the categories proposed by other scholars which have been deemed to be the manifestation of the Fundamental categories 'Matter' are eventually of three kinds though they have been called by different names. They are:

(i) Material/Substance;

(ii) Properties; and

(iii) Processes/Methods.

This is in conformity with the concept of 'Matter' as postulated by Ranganathan, which consists of 'Matter Material'; 'Matter Property'; and 'Matter Method'.

Similarly, the categories reducible to Fundamental Category 'Energy' have been denoted by the term 'Action'; 'Operation'; and 'Activities' by the scholars. This is why Ingetrant Dahlberg has rightly remarked that:

> If one considers Classification from a purely methodological point of view, it does indeed appear that Ranganathan's ideas and contributions have thus for not been replaced by any better ones. In fact, they have not as yet been discussed every where, and there has been little movement towards their adoption throughout the world. (Dahlberg 1977: 43)

Coming to the semantic principles propounded by different scholars such as Cutter, Kaiser, Austin,

Bhattacharyya, Taube, Mooers, Perry Kent and Rees, Aitchison and Garfield, etc., they can be grouped under three categories: Equivalence, hierarchical and associative as shown in the beginning of this chapter.

If these principles are analysed at seminal level, it can be easily recognised that these are comprehensively contained in the following canons, principles and postulates of Ranganathan.

(1) Canons for Characteristics;
(2) Canons for Array;
(3) Canons for Chain;
(4) Canons for Filiatory Sequence;
(5) Principles for Isolate Sequence; and
(6) Canons for Verbal Plane.

Equivalence Relations: Canons for Verbal Plane.

Heirarchical Relations:

(1) Canons for Characteristics;
(2) Canons for Array;
(3) Canons for Chain;
(4) Canons for Filiatory Sequence; and
(5) Principles for Isolate Sequence.

Associative Relations: Facet relationship.

Thus we can say -- with confidence that the Ranganathan model has all the Germs of Seminal ideas contained with to guide further work in the field of information retrieval. S.D. Niel, while explaining the boundaries of information science, has aptly remarked that:

> S.R. Ranganathan who invented much of the language we use to talk about the theory of classification, recognised the oral nature of the universe of knowledge. He saw it as 'multi-dimensional' he knew he faced the insolvable problem of maping multi-dimensional space into one linear dimension. The multiple and simultaneous relationships of the world of knowledge were well understood by Ranganathan, whose roots were in the oral tradition of India, but he was also a person who had an extremely scientific attitude and who suffered sleepless nights and nightmares as he struggled with the maddening problem. (Niel 1992: 152).

CITED REFERENCES

Aitchison, J. "The Thesaurofacet: A multipurpose retrieval language tool". Journal of Documentation, 26: 1970.

Aitchison, J. "Thesaurofacet - A new concept in Subject Retrieval in the Seventies". In: New Directions Subject Retrieval in the Seventies by Hans Wellisch and Thomas D. Wilson, Maryland; University of Maryland, 1972.

Aptagiri, Devika V. et al. "A Frame Based Knowledge Representation Paradigm for Automating POPSI", Knowledge Organization, 22(3/4), 1995.

Austin, D.W. "Prospect for a New General Classification", Journal of Library Science, 1(3), 1969.

Austin, Derek. "CRG research into a freely faceted scheme". In: Classification in the 1970's: A Second Look, Ed. by Author Maltby. London: Clive Bingley, 1976.

Austin, D. "The PRECIS System for Computer Generated Index and its Use in the British National Bibliography". In: Subject Retrieval in the Seventies New Directions, ed. by Hans Wellisch and T.D. Wilson. Maryland: School of Library and Information Science, 1972.

Austin, D. "Basic concept classes and primitive relations". In Universal Classification 1 Subject Analysis and Ordering Systems, ed. by I. Dahlberg, 1982. Frankfurt: Indek's Verlag, 1982.

Austin, D. PRECIS: A Manual of Concept Analysis and Subject Indexing. London: The British Library Bibliographic Services Division, 1984.

Bhattacharyya, G. "Subject indexing language: Its theory and practice". DRTC Refresher Seminar 13; 1981.

Bhattacharyya, G. "POPSI: A source language for Organizing and Associative Classification." Library Science 19, 1982.

Bliss, H.E. The Organization of Knowledge and the System of Sciences, New York: Halt, 1929.

Bliss, H.E. The Organization of Knowledge in Libraries and the Subject Approach to Books. New York: The H.W. Wilson & Co., 1933.

Bliss, H.E. A Bibliographic Classification, Extended by Systematic Auxiliary Schedules for Composite Specification and Notation. New York: Wilson, 1952-53.

Brown, J.D. Subject Classification for the Arrangement of Libraries and the Organization of Information, With Tables Indexes etc. for the Subdivision

of Subjects. 3rd ed. Revised and Enlarged by J.D. Stewart. London: Grafton & Co., 1939.

Classification Research Group. "Proposal for a New Classification." In Some Problems of General Classification Scheme: Report of Conference held in London, January 1963. Library Association, 1964.

Coates, E.J. Subject Catalogues: Heading and Structure. London: The Library Association, 1960.

Cordonnier, G. "Classification et Classement". Bulletin d' Information Scientifique et Technique. 6; 1944.

Cordonnier, G. "Optim de I'organization documentaire". Revue de la Documentation, 27(7), 1960.

Cyrstal, D. A First Dictionary of Linguistics and Phonetics. London: Andre Deutsch, 1980.

Cutter, C.A. Rules for a Dictionary Catalogue, 4th Ed. Rewritten. Washington: Government Printing Office, 1904 (First Published in 1876).

Dahlberg, I. "Major Development in Classification". Advances in Librarianship, 7; 1977.

Farradane, J.E.L. "Fundamental Fallacies and Now Need in Classification". In Sayers Memorial Volume, Essays in Librarianship in the Memory of William Charles Birwick Sayers, ed. by D.J. Foskett and B.I. Palmer, 1961.

Foskett, D.J. "Documentation in Occupational Safety and Health". Reveu de la Documentation 27(3), 1960.

Foskett, D.J. and Pamer, B.I. "Classification and Integrative Levels." In Sayares Memorial Volume. London: The Library Association, 1961.

Foskett, D.J. Classification for a General Index Language. London: The Library Association, 1970.

Gardin, J.C. SYNTOL Rutgers Series on Systems for Intellectual Organization of Information, ed. by Susan Artandi Vol. II, Rutgers, New Brunswick, The State University, 1965.

Garfield, E. "Citation Indexes for Science". Science 122, 1955.

Garfield, E. "Science Citation Index: A New Dimension in Indexing". Science 144, 1965.

Grolier, Eric de. A Study of General Categories Applicable to Classification and Coding in Documentation. Paris: UNESCO, 1962.

Groupe d' Etudes Documentaires pour les Sciences Humaine of the Association Marc Bloch Report I, page I.

Kaiser, J.O. Systematic Indexing. London: Pitman, 1911. (The Card System Series) Vol. II.

Kyle, Barbara. "Toward a Classification for Social Science Literature". American Document, V. 9, 1958.

Library Association Record. "The Need for a Faceted Classification as the Basis of all Methods of Information Retrieval." Library Association Record 57:7, 1955.

Luhn, H.P. "Keyword-in-Context for Technical Literature (KWIC Index)." American Doc. 11; 1960.

Luhn, H.P. (1952-64). "Selected Works". In H.P. Luhn: Pioneer of Information Sciences, Ed. by C.K. Schultz. New York: Spartox, 1968.

Mooers, Calvin M. "A Mathematical Theory of Language Symbols in Retrieval". In International Conference on Scientific Information, Vol. II. Washington: National Academy of Science, 1959.

Mooers, Calvin M. The Indexing Language of an Information Retrieval System. Papers presented at an Institute Conducted by the Library School and the Centre for Continuation Studies. ed. by Wisley Simonton. Minneapolis Minn; the Centre, 1963.

Needham, C.D. Organizing Knowledge in Libraries: An Introduction to Information Retrieval. 2nd Rev. Ed. London: Andre Deutsch, 1971.

Niel, S.D. Dilemmas in the Study of Information. New York: Greenwood Press, 1992.

Perry, J.W., Kent, A. and Berry, M.M. Machine Literature Searching. New York: Interscience Pubs., 1956.

Perry, J.W. and Kent, A. Documentation and Information Retrieval. cleveland: Western Reserve University, 1957.

Perry, J.W. et al. Tools for Machine Literature Searching. New York: Interscience Pubs., 1958.

Ranganathan, S.R. Colon Classification, 6th ed. Madras Library Association. London: Asia Publishing House, 1960.

Ranganathan, S.R. A Descriptive Account of Colon Classification. Bombay: Asia Publishing House, 1965.

Ranganathan, S.R. Prolegomena to Library Classification. 3rd Ed. Bangalore: Sarda Ranganathan Endowment for Library Science, 1967.

Taube, M. Studies in Coordinate Indexing. 4 Vols. Washington: Documentation Incorporated, 1953-57.

Taube, M. "Notes on the Use Roles and Links in Coordinate Indexing". American Documentation 12(2); 1961.

Vickery, B.C. Faceted Classification: A Guide to Construction and Use of Special Scheme. London: Aslib, 1960.

Vickery, B.C. Classification and Indexing in Science. 3rd Ed. London: Butterworths, 1975.

Vickery, B.C. "Knowledge Representation: A Brief Review". Journal of Documentation, 42(3); 1986.

CHAPTER IV

APPLICATION OF LINGUISTIC ANALYSIS IN INFORMATION RETRIEVAL TOOLS

APPLICATION OF LINGUISTIC ANALYSIS IN INFORMATION RETRIEVAL TOOLS

4.0. Introduction

Since all the semantic and syntactic principles propounded by different scholars are contained in the canons, postulates and principles of Ranganathan, it was considered advisable at the next stage as how for these canons, postulates and principles have been applied in information retrieval tools. It was decided to select two information retrieval tools. The criteria of Pre-Coordinate and Post-Coordinate indexing helped in selecting the tools. One was selected 'PRECIS' (Preserved Context Indexing System), which belongs to the category of Pre-coordinate indexing and the other one selected was "The UNESCO:IBE Education Thesaurus" which belongs to the category of Post-Coordinate indexing. The reason behind the selection of these two tools is that the PRECIS is one of the latest indexing tools which has world wide recognition. It has been adopted by different National Bibliographies for the preparation of their index. These includes: BNB; Australian National Bibliography; Deutsche

Bibliothek in Frunkfurt and National Film Board of Canada. The UNESCO:IBE Education Thesaurus was another choice, because this tool is prepared by a group of well recognised world wide experts engaged in the field of Education. Now we will try to explore, how far Ranganathan's conceptual model is successful in regard to these two tools of information retrieval.

4.1. PRECIS (Preserved Context Indexing System)

4.11. Syntax

The syntax of PRECIS is based on Ranganathan's analytico-synthetic approach which assumes that a subject, however complex, can be broken down into its separate components or facets, and that they can be reorganized consistently into a standard pattern by reference to a general decision-making model. The decision making model of Ranganathan is 'Generalized Facet Structure' supported by postulates and principles. It reads as follows:

(BS)

[1P1], [1P2], ...,, [1PLp]

[1M1]; [1M2]; ....; [1MLm]

[1E]:

[2P1],[2P2]; ....,[2PLq]

[2P1]; [2PM2]; .... ;[2MLm]

[2E]:

... ...

[RP1], [RP2], ... ,[RPLr]

[RM1]; [RM2]; .... ;[RMLr]

[RE]:

[SL].[S2]. .... .[SLs]

[T1]' [T2] .... [TLt]

Where Lp, Lm, Lq, Ln, Lr, Ls, Lt may have any integral value. (Ranganathan 1967:464)

Whereas the decision making model of PRECIS consists of the following schema of role operator:

PRECIS: Schema of Role Operators

Primary Operators

Environment of Core Concepts	O Location
Core concepts	1. Key system Things when action not present Things towards which an action is directed, e.g. Object of Transitive action, performer of intransitive action. 2. Action: Effect on action. 3. Performer of transitive action (agent, instrument) Intake: Factor.

Extra core concepts	4.	Viewpoint-as-form: Aspect
	5.	Selected instance, e.g., Study region, sample population.
	6.	Form of document: Target user.

Secondary operators

Coordinate concepts	f	Bound coordinate concept.
	g	Standard coordinate concept.
Dependent elements	p	Part: Property
	q	Member of quasi generic group.
	r	Assembly
Special classes of action	s	Role definer, Directional property.
	t	Author attributed association.
	u	Two way interaction.

(Austin 1984: 307)

If we analyse these role operators at seminal level, it can be derived that the schema of Role Operators is based on 'General Facet Structure' of Ranganathan. There is an apparent equivalence between the primary operators and some of the secondary operators of PRECIS and the Fundamental Categories of Ranganathan. Rest of the operators of PRECIS are derivatives of some other

concepts propounded by Ranganathan. More explanation of this is as follows:

Environment

The primary core role operator of PRECIS 'Environment' can be deemed to be the manifestation of the Fundamental Category 'Space' in Ranganathan.

For example: Library management in Canada.

In the present example, Canada is shown by PRECIS as Environment whereas Ranganathan takes it as manifestation of the Fundamental Category = 'Space'.

Key System

Key system of PRECIS includes the following concepts:

	Classes of concepts	Examples
1.	Physical countable things	Penguins, Telephones
2.	Physical non-countable things (materials)	Timber, Granite
3.	Abstract entities	Heat; time
4.	System of abstract entities	Christianity; Law.
5.	System of physical and abstract entities	Primary schools; Anglican churches, Health services.
6.	Organs of the body	Nose, Lungs, Respiratory System, etc.

These concepts have been covered by Ranganathan under the Fundamental Category 'Personality'.

Action

This operator includes two types of action. These are: transitive action; and intransitive action. In the subject statement, 'Harvesting apples' is an example of transitive verb. On the other hand, 'Camping' in the subject statement 'Camping by boys' is intransitive action. Ranganathan considers these actions as manifestation of the Fundamental Category 'Energy'.

Performer of the transitive action (agent, instrument) Intake; Factor

This operator stands for agents, instruments intake, factors, tool etc. The following examples will illustrate it:

1. 'Frost damage to building'

Here 'frost' is the agent for the action damage.

This is covered by Ranganathan under the Fundamental Category Personality in its second Round.

2. Hunting of rodents by foxes.

In this example, foxes are the performer of the action hunting.

This is covered by Ranganathan's Fundamental Category Personality.

Similarly, extra core role operators 4, 5 and 6. Viewpoint-as-form aspect, selected instance, Form of document respectively are near to the concept of Ranganathan's Common Isolates. The following examples provided in PRECIS Manual may suffice to support it.

1. 'Christian viewpoints on marriage'.

Here viewpoint according to Ranganathan is a Posteriorising Common Isolate assisted by 'g' and can be amplified by Christian.

2. 'Visitor guide to Tower of London'

In this example, Visitors guide of PRECIS is again Anteriorising Common Isolate.

Coming to the secondary operators, under the Group Coordinate concepts and some of the concepts under the group Dependent elements it is apparent that 'Bound' coordinate concepts; standard coordinate concepts; concepts denoting to member of quasi-generic group and Assembly concepts are semantic operators, covered by Ranganathan under Canons For Array and Chain.

Further role operators 'part' and 'property' manifesting of the Fundamental Category of 'Personality' and 'Matter' respectively designed by Ranganathan.

Moving down to the group of secondary operators under the label 'Special Classes of Action', it may be easily ascertained that the operators like 'Author attributed association' and 'two-way interaction' are derivatives of the concept of Phase-Relation introduced by Ranganathan. Some of the examples given below will illustrate the above statement.

1. Foreign relations between Brazil and Portugal.

It is an example given in PRECIS manual under 'two way interaction' operator. Ranganathan has covered it under Intra-Facet Phase-Relation.

2. Economic development in Brazil compared with economic development in France.

It is an example given in PRECIS manual under 'Authro-attributed association operator'. Ranganathan has covered it under 'Intra-Facet Phase-Relation.'

3. Mathematics for navigators.

It is an example given in PRECIS manual under

'Target user' operator. Ranganathan has covered it under Intra.Subject Phase Relation.

'Role operator', 'Role definer' is close to 'Energy' concept of Ranganathan. Some of the examples provided in PRECIS Manual under this operator illustrate the above statement.

1. Use of psychotropic drugs in behaviour modification in animals.
2. Software package for microcomputer systems applied to school administration.

4.12. Semantics of PRECIS

Thesaurus contains the semantics of PRECIS. Terms have been connected in the Thesaurus by three basic relationship:

" - The equivalence relationship;
- The hierarchical relationship; and
- The associative relationship."

(Austin 1984: 200)

(i) Equivalence Relationship

The relationship that links terms referring to the same concept. Thus these are making an equivalence set. To avoid confusion same mechanism should be applied

so that these terms can be properly indexed. For this purpose preferred and non-preferred terms have been used. For example, Hypersensitivity and Allergy.

In this, Allergy is a preferred term as common used term, whereas Hypersensitivity is a scientific and unpopular term. Therefore, the sequence will be:

Allergy; Hypersensitivity.

The equivalence relationship covers two classes of terms, known as 'Synonyms' and 'Quasi-Synonyms'.

Synonyms

Various kinds of synonyms are found. Some of the synonyms are as follows:

(a) Terms of different linguistic origin, e.g., Polyglot dictionaries; Multlingual dictionaries.

(b) Popular names and scientific names, e.g., Penguin; Sphenisciformes.

(c) Common names and trade names, e.g., Ball point pens; Biros (trade name). Preference should be for common names.

(d) Current or Favoured terms versus outdated or deprecated terms, e.g., Chairperson; Chairmen

Radio; Wireless

San; Bushmen.

(e) Terms originating from different culture, e.g.,

Mobile libraries; Bookmobiles, etc.

In all these cases only preferred term should be selected for entry element.

Quasi-Synonyms

Quasi-synonyms are the terms whose meaning are generally recognised as different in common usage, yet they are treated as synonyms for indexing purposes (Austin 1984: 202). For example,

Wetness; Dryness

Hardness; Softness.

(ii) The Hierarchical Relationship

Hierarchical relationship is based on the notions of super-ordinate and subordinate, where the Superordinate term represents whole or a complete class, and the subordinate indicates part or constituent or its member. This relationship covers three sub-types:

(1) The Gensic relationship, e.g.

Documents and Books

(Genus ..... species)

2. The hierarchical whole-part relationship, e.g.,

   World,
   Asia,
   India,
   Meghalaya,
   Shillong.

3. The instance relationship links a general class of things or action, expressed by a common noun, with a particular instances of that class expressed by a proper name. (Austin 1984: 206), e.g.,

   Man - General class

   Mohan, Ram - specific instances.

(iii) The Associative Relationship

It covers the link between a pair of terms that are not members of equivalence set, nor can they be organised as a hierarchy where one of the terms is subordinate to the other, yet they are mentally associated to such an extent that we can reasonably assume that a user searching the index under one of the terms would be helped by a reference to entries containing the other. For example, 'Birds' and 'Ornithology'.

Two kinds of concept can be linked in associative relationship:

- Those belonging to the same categories e.g., Plants and Herbicides; 'Ships' & 'Boats'.

- Those belonging to different categories, e.g.:

(a) A discipline or field of study and the object or phenomena studied, e.g.:

Birds; Ornithology

(b) An action and production of that action, e.g.:

Weaving; Cloth

Welding; Welds.

(c) An operation or process and its agent or instrument, e.g.:

Typing; Typewriter.

(d) An action and its patient, e.g.:

Harvesting; Crops

Imprisonment; Prisoners.

(e) Concepts related to their properties, e.g.:

Poisons; Toxicity

Perception; Acuity

(f) Concepts and their counter operations or agents, e.g.:

Pain; Analgesia; Analgesics

Aircraft; Anti-aircraft weapons, etc. (Austin 1984: 209).

Basic relations and the directions of references in printed indexes:

(1) Equivalence relationship - can be printed in one direction, i.e., from non-preferred term to its preferred terms, e.g.:

Sphenisciformes, See Penguin.

(2) Hierarchical relationship can be expressed two directions, e.g.:

Birds	Penguins
See also	See also
Penguin	Birds.

(3) Associative relationship: See also reference should be made in all cases:

Pain	Aircraft
See also	See also
Analgesia	Model Aircraft, etc.

As already shown, these relationships are derivatives of Ranganathan's canons for Array, Canons for Chain, Canons for Verbal Plane and Principles of Isolate Sequence.

4.2. The UNESCO : IBE Education Thesaurus

Since the start of computerised information retrieval systems based on Post Coordination, thesaurus has become the most potential tool of information today.

The thesaurus is the latest type of indexing language to be developed. It appeared in the late 1950s,

designed for use with the emerging post coordinate indexing systems. It differs from a mere list of terms, being fully organised, so the a priori relationships between concepts are made explicit. For the recognition of thesaurus as world wide information retrieval tool, it was felt necessary to form international standards to meet the requirement in the preparation of these thesaurus. In this direction several attempts have been made and ISO 2788: 1986 an international standard for mono lingual and ISO 5964: 1985 for multilingual standard are results of these efforts.

The features of the thesaurus covered by these standards are vocabulary control, thesaurus relationships and thesaurus display.

4.21. Vocabulary Control

Concepts are represented by indexing terms which are either preferred or non-preferred terms. The preferred terms are used for indexing and searching and the non-preferred terms functions as lead-ins to the preferred terms. If the thesaurus is used for searching a natural language system, both preferred and non-preferred terms may be used as search terms.

Rules for covering the form of the terms are the use of noun or noun form, choice of singular or plural,

treatment of abbreviations, scientific or popular names, slang words or trade names.

To avoid homographs and clarification of meaning, scope note is provided to treat the compound terms. Three rules for factorising of compound terms are prescribed by standards. These are:

(1) When a property or part is qualified by its containing whole, for example:
Concrete strengh, should be factored as the property.
Strength, qualified by the containing whole, concrete.

(2) When a transitive action is modified by the object of the action as in Aluminium Welding which should be split into Aluminium + Welding.

(3) An intransitive action should not be modified by the name of the performer or the action. In Infant Behaviour, the intransitive action, Behaviour, is qualified by the performer of the action, Infant and should be factored into Infant + Behaviour.

4.22. Thesaurus structure and relationship

A thesaurus believes in displaying semantic relationships among the terms and leaves syntax to the care of the users of the system. The user is expected to develop his own syntax by applying techniques like Boolean operators etc. Generally a thesaurus displays three types of semantic relationship among the terms. These are:

1. Equivalence;
2. Hierarchical; and
3. Associative.

Equivalence is represented by the device 'USE FOR' or USE; Heirarchical is shown by 'BT' or 'NT' and Associative by 'RT'.

4.221. Equivalence Relationships

USE/UF relationships include:

(a) Synonyms, for example: ANTENNAS

UF Aerials

Aerials

use ANTENNAS

(b) Quasi synonyms, for example:

Hardness

UF Softness

Softness

use Hardness.

4.222. Hierarchical Relationships

These include Broader term and Narrower term relationships, for example:

Capital Markets / BT Financial markets.

Financial Markets / NT Capital markets.

There are three forms of heirarchical relationships. These are:

(a) Generic relationship: This identifies the relationship between a class and its members species. For example:

BIRDS/NT Parrots

PARROTS/BT Birds.

(b) Whole part relationship: In this category four types of conditions are prescribed by standards. Other than these four conditions, rest are treated as associative relationship. These four conditions are:

(i) Geographical divisions;

(ii) Disciplines and their divisions;

(iii) Hierarchical serial structure; and

(iv) Systems and organs of the body.

For example:

Organs of the body - Gastrointestinal systems
BT Digestive systems
NT Intestine
NT Stomach.

(c) Polyhierarchical Relationship: This occurs when a concept belongs to more than one category. For example:

PRINTING EQUIPMENTS	COMPUTER PERIPHERAL EQUIPMENT
NT Computer printers	NT Computer printers
Computer Printers	
BT Computer peripheral equipment	
BT Printing equipment.	

4.223. Associative Relationship

The associative relation is usually employed to cover those relations between concepts that are neither heirarchical nor equivalent. (e.g. Similarly antonymity) (UNESCO UNISIST 1973: 73). For example:

TEACHING	BAKING
RT TEACHING AIDS	RT BREAD
EDUCATION	BREAD
RT TEACHING	RT BAKING

4.23. Thesaurus Displays

The standard recognises two basic types of thesaurus displays:

(i) Alphabetical,

(ii) Systematic.

(a) Alphabetical Displays; This Alphabetical thesaurus displays term/descriptions arranged in alphabetical order. Each term is supported by the following relationships:

SN - Scope Note

UF - Used for

BT - Broader term

NT - Narrower term

RT - Related term

(b) Systematic Displays: A Systematic thesaurus as defined by the standard is divided into two parts, the one including categories or hierarchies of terms arranged in subject order, and the other an alphabetical section, which direct the user to the appropriate parts in the systematic section. The link between them is the notation. The systematic display is helpful both for indexer and searcher. For example:

VOCATIONAL TRAINING

UF: JOB TRAINING

BT: TRAINING

VOCATIONAL EDUCATION

NT: FULL TIME TRAINING

OFF THE JOB TRAINING

ON THE JOB TRAINING

PART TIME TRAINING

TRAINING BY STAGES

RT: 351

4.24. IBE: Education Thesarus

The basic framework of IBE: Education Thesaurus is also not different from other Thesauri. It has one added feature: Faceted Array of Descriptors and Identifiers.

The following examples will suffice to show the structure of this thesaurus:

(1) Educational Programmes

SN: Series of planned activities, with administrative connotation relating to development of institution, what goes on in them and budget provisions; to be distinguished from instructional programmes.

BT: Programmes.

NT: Compensatory Education Programmes

Exchange Programmes

Guidance Programmes

Liteerary Campaign

RT: 271

RT 271 refers a user to the Faceted array of descriptors and identifiers, where he gets the related terms as follows:

271 EDUCATIONAL PROGRAMMES

COMPENSATORY EDUCATION PROGRAMMES

EDUCATIONAL PROGRAMMES

EMERGENCY PROGRAMMES

EXCHANGE PROGRAMMES

GUIDANCE PROGRAMMES

LITERARY PROGRAMMES

PROVINCIAL PROGRAMMES

RECREATIONAL PROGRAMMES

REHABILITATION PROGRAMMES (P. 72)

(2) School System

UF: FORMAL EDUCATION

SN: RELATING PARTICULARLY TO STRUCTURAL ASPECTS OF INSTITUTIONS OF FORMAL EDUCATION

BT: EDUCATIONAL ORGANIZATION

NT: PUBLIC SCHOOL SYSTEMS

RURAL SCHOOL SYSTEMS

RT: 260

RT 260 refers to users to the faceted array of descriptors and identifiers for the related terms:

260 SCHOOL SYSTEM
EDUCATIONAL ORGANIZATION
PUBLIC SCHOOL SYSTEMS
RURAL SCHOOL SYSTEMS
SCHOOL SYSTEMS (P. 163)

(3) Salaries

UF: REMUNERATION
SALARY SCALE
WAGES

BT: INCOME

NT: TEACHER SALARIES

RT: 242

RT 242 refers to user to the faceted array of descriptors and identifiers for related terms.

242 SALARIES
FELLOWSHIPS
PREMIUM PAY

SALARIES

SCHOLARSHIPS

STUDENTS LOANS

TEACHER SALARIES

TRAINING ALLOWANCES (P. 159)

(4) Educational Administration

SN: RELATING TO PART OR ALL OF EDUCATIONAL SYSTEM

BT: ADMINISTRATION

NT: COLLEGE ADMINISTRATION

SCHOOL ADMINISTRATION

RT: ADMINISTRATIVE ORGANIZATION

200

RT 200 refers to users for related term in the facet of array of descriptors and identifiers.

200 ADMINISTRATIVE POLICIES

ADMINISTRATION

ADMINISTRATIVE POLICIES

ADMINISTRATIVE PRINCIPLES

ADOPTION OF INNOVATIONS

CENTRALIZATION

DECENTRALIZATION

EDUCATIONAL ADMINISTRATION

EDUCATIONAL AIMS
EDUCATION COORDINATION
EDUCATION INNOVATIONS
PROMOTION POLICIES
STUDENT WELFARE
TRANSFER POLICIES (P. 67)

(5) Teaching Material

SN: Patterns of teacher behaviour that are recurrent applicable to various subject matters, characteristics of more than one teacher and relevant to learning.
BT: EDUCATIONAL METHODS
NT: ACTIVITY METHODS
AUDIOLINGUAL METHODS
DEDUCTIVE METHODS
INDUCTIVE METHODS
RT: 355

RT 355 refers to user for related terms in the faceted array of descriptors and identifiers.

355 EDUCATIONAL METHODS
ACTIVITY METHODS
AUDIOLINGUAL METHODS

CLASSROOM COMMUNICATION

DEDUCTIVE METHODS

INDUCTIVE METHODS

INTERVENTION

LECTURES

SELF INSTRUCTIONAL METHODS

SEMINARS

SIMULATION

TEACHING METHODS

TUTORING (P. 186)

Analysing the relations shown by devices 'USE', 'BT', 'NT' at seminal level, the present study has already shown in chapter three that Ranganathan has comprehensively covered these relations in his conceptual framework of Canons for Verbal Plane, Canons for Chain, Canons for Array and Principles of Isolate Sequence.

Associative relationships which include among others, concepts related to their processes and properties, operations and their agents or instruments, an action or the product of the action, the whole-part relationship other than the heirarchical whole-part, and many others, are well covered by Ranganathan under facet-relationship.

Thus the theory of indexing language formulated by Ranganathan is valid even in computerised information retrieval systems at semantic level. However, syntactic part of his theory is not applicable to post coordinate indexing languages. These languages believe in combination of concepts of the search stage. The devices used for combination are many. Some of the often used devices are:

1. Boolean search logic
2. Weighted-term search logic
3. Truncation
4. Adjacency and proximity operators
5. Field level searching
6. Rang and limiting operators
7. Free text searching
8. Best match searching based on structural analysis.

Application of these devices on mass-scale has brought into focus the problem of 'noise' in retrieval. The recent researches are concentrating on developing mechanisms and techniques which may lead to minimizing of 'noise'. The area is still at explorating stage. However, knowledge-based techniques from the field of

Artificial Intelligence are appearing to show rays of hope. If study in depth is undertaken of the knowledge-based techniques vis-a-vis the conceptual framework of Ranganathan, it may appear that Ranganathan was ahead of his time.

CITED REFERENCES

Austin, D. PRECIS: A Manual of Concept Analysis and Subject Indexing, 2nd ed. London: The British Library Bibliographic Services Division, 1984.

International Organization for Standard ISO 2788. Guidelines for the Establishment and Development for Monolingual Thesauri, 2nd ed., Geneva ISO, 1986.

International Organization for Standard ISO 5964. Guidelines for the Establishment and the Development of Multilingual Thesauri, Geneva, ISO, 1985.

Ranganathan, S.R. Prolegomena to Library Classification, Bangalore: Sarda Ranganathan Endowment for Library Science, 1967.

The UNESCO: IBE Education Thesaurus, 2nd ed., Paris: The UNESCO Press, 1975.

CHAPTER V

CONCLUSIONS AND FINDINGS

CONCLUSIONS AND FINDINGS

5.0. Conclusions and Findings

5.1. Conclusion

The study presented in the thesis revolves around the three hypotheses assumed in the first chapter. These hypotheses are, namely:

(i) Whether there is a correspondence between Linguistic Analysis and Information Retrieval at only two levels of Linguistic Analysis, i.e., Semantics and Syntax;

(ii) Whether principles of Information Retrieval propounded by different scholars at semantic and syntactic levels bear upon the Canons, Principles and Postulates developed by Ranganathan; and

(iii) Whether the conceptual framework developed by Ranganathan for syntax is loosing its grip in the field of information retrieval.

The first hypothesis stands verified in chapter two. It has been established in this chapter that there

is correspondence between Linguistics and Information Retrieval. It was also ascertained in this chapter that there are five types of linguistic analysis, viz.

1. Segmental Analysis;
2. Componential Analysis;
3. Distributional Analysis;
4. Discourse Analysis; and
5. Immediate Constituent Analysis.

These five types of Linguistic Analysis are used at three levels, i.e., Phonology, Syntax and Semantics. An examination of these techniques and levels of Linguistic Analysis in context of indexing languages revealed that an indexing language functions at two levels, i.e., Semantics and Syntax. It has also been observed in this chapter that the Phonological level, missing from the general indexing languages, is now finding a place in retrieval tools being developed on speech recognition techniques. The second hypothesis led to a survey of the principles propounded by different scholars in the field of Information Retrieval. Only those principles were taken into consideration which had been developed on the basis of Linguistic Analysis and which worked at either semantic level or syntactic level or both. This survey led to the following findings:

(1) The principles laid down by different scholars at semantic and syntactic level are either based on or reducible to the principles propounded by Ranganathan.

(2) Conceptual model of Ranganathan, consisting of conons, postulates and principles has all the germs of seminal ideas contained in it to guide further researches in the field of Information Retrieval. In fact, his seminal ideas have not been fully exploited. Even the ideas expressed by the latest thinkers in the field like Dahlberg and Fugman are more or less prototypes of Ranganathan's ideas couched in different expressions.

Thus second hypothesis is proved true and its verification details have been provided in Chapter III.

A test of Ranganathan's conceptual model was undertaken in the 4th chapter. Two Information Retrieval Tools, one belonging to Pre-coordinate Indexing System and other one belonging to Post-coordinate Indexing System were chosen. The choice fell on PRECIS (Preserved Context Indexing System) and The UNESCO:IBE Education Thesaurus. A study of these tools revealed that semantic principles of Ranganathan are still valid in both Pre- and Post-Coordinate Indexing systems. On the other hand, the Syntactical principles of Ranganathan are no more applicable

in context of Post-Coordinate Indexing Systems. These systems are relying on other techniques, which are:

(i) Boolean search logic;

(ii) Weighted-term search logic;

(iii) Turncation;

(iv) Adjacency and proximity operators;

(v) Field level searching;

(vi) Free text searching;

(vii) Rang and limiting operators; and

(viii) Best-match searching based on structural analysis.

Thus the third hypothesis is also proved to be true.

5.2. Suggestions

No doubt, the new techniques of Post-Coordinate Indexing System are very potential and playing vital role in machine retrieval. However, the phenomenon of 'noise' produced by these techniques is quite disturbing. The scholars are, therefore, now turning to knowledge-based systems by applying tools and techniques of knowledge representation developing in the field of Artificial Intelligence. The results are still not conclusive.

However, one thing is very clear that Ranganathan's conceptual model has enough seminal ideas contained in it to give a direction to the knowledge-based systems. Research needs to be taken on a large scale in this area in future.

Whatever form future trend may take, the ultimate final stage of alleged omniscience is never feasible. The following words of Xenophanes, translated by Karl Popper in this context make the point:

> The God did not reveal, from the beginnings,
> All things to us, but in the course of time
> Through seeking we may learn and know things better.
> But as for certain truth, no man has known it,
> Nor shall he know it, neither of the Gods
> Nor yet of all the things of which I speak.
> For even if by chance he were to utter
> The final truth, he would himself not know it
> For all is but a woven web of guesses.
>
> (Karl Popper 1972: 152-3)

CITED REFERENCES

Popper, Karl. Conjectures and Refutations, 4th ed., London: Routledge and Kegan Paul, 1972.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aitchison, J. "The Thesaurofacet: A Multipurpose Retrieval Language Tool." Journal of Documentation, 26:1970.

Aitchison, J. "Thesaurofacet - A New Concept in Subject Retrieval in the Seventies." In New Directions Subject Retrieval in the Seventies, by Hans Wellisch and Thomas D. Wilson. Maryland: University of Maryland, 1972.

Aptagiri, Devika V. et al. "A Frame Based Knowledge Representation Paradigm for Automating POPSI". Knowledge Organization 22(3/4): 1995.

Austin, D.W. "Prospect for a New General Classification". Journal of Library Science 1(3): 1969.

Austin, D.W. "CRG Research into a Freely Faceted Scheme." In Classification in the 1970's: A Second Look. Ed. by Author Maltby. London: Clive Bingley, 1976.

Austin, D.W. "The PRECIS System for Computer Generated Index and its Use in the British National Bibliography". In Subject Retrieval in the Seventies New Directions. Ed. by Hans Wellisch and T.D. Wilson. Maryland: School of Library and Information Science, 1972.

Austin, D. "Basic Concept Classes and Primitive Relations." In Universal Classification I Subject Analysis and Ordering Systems, Ed. by I. Dahlberg, 1982, Frunkfurt: Indek's Verlag, 1982.

Austin, D. PRECIS: A Manual of Concept Analysis and Subject Indexing. London: The British Library Bibliographic Servicees Division, 1984.

Belkin, N.J. "Anomalous States of Knowledge as a Basis for Information Retrieval." Canadian Journal of Information Science V. 6, 1980.

Bhattacharyya, G. "Subject Indexing Language: Its Theory and Practice." DRTC Refresher Seminar 13, 1981, BA.

Bhattacharyya, G. "POPSI: A Source Language for Organizing and Associative Classification." Library Science 19, 1982.

Bhattacharyya, G. "Role of Classification and Indexing in Information Retrieval." DRTC Annual Seminar V. 20, 1983.

Bliss, H.E. The Organization of Knowledge and the System of Sciences. New York: Halt, 1929.

Bliss, H.E. The Organization of Knowledge in Libraries and the Subject Approach to Books. New York: The H.W. Wilson & Co., 1933.

Bliss, H.E. A Bibliographic Classification, Extended by Systematic Auxiliary Schedules for Composite Specification and Notation. New York: Wilson, 1952-53.

Brown, J.D. Subject Classification for the Arrangement of Libraries and the Organization of Information, With Tables, Indexes etc. for the Subdivision of Subjects, 3rd ed. Rev. and Enlarged by J.D. Stewart. London: Grafton & Co., 1939.

Chomsky, Noam. Syntactic Structure. The Hageu: Mouton Pub., 1957.

Classification Research Group. "Proposal for a New Classification." In Some Problems of General Classification Scheme: Report of Conference held in London, Jan. 1963. Library Association, 1964.

Coates, E.J. Subject Catalogues: Heading and Structure. London: The Library Association, 1960.

Cordonnier, G. "Classification et Classement". Bulletin d' Information Scientifique et Technique, 6; 1944.

Cordonnier, G. "Optim de I'organization documentaire." Revue de la Documentation 27(7), 1960.

Crystal, D. A First Dictionary of Linguistics and Phonetics. London: Andre Deutsch, 1980.

Cutter, C.A. Rules for a Dictionary Catalogue. 4th Ed. Rewritten. Washington: Government Printing Office, 1904 (First Published in 1876).

Dahlberg, I. "Major Development in Classification." Advances in Librarianship 7; 1977.

Devy, G.N. After Amnesia: Traditional and Changes in Indian Literature Criticism. Hyderabad: Orient Longman, 1992.

Farradane, J.E.L. "Fundamental Fallacies and New Need in Classification." In Sayers Memorial Volume, Essays in Librarianship in the Memory of William Charles Birwick Sayers. Ed. by D.J. Foskett and B.I. Palmer. London: Library Association, 1961.

Farradane, J.E.L. "Analysis and Organization of Knowledge for Retrieval." ASLIB Proceeding 22(2); 1970.

Foskett, A.C. The Subject Approach to Information, 4th ed. London: Clive Bingley, 1982.

Foskett, D.J. "Documentation in Occupational Safety and Health." Revue de la Documentation 27(3); 1960.

Foskett, D.J. and Palmer, B.I. "Classification and Integrative Levels." In Sayers Memorial Volume. London: The Library Association, 1961.

Foskett, D.J. Classification for a General Index Language. London: The Library Association, 1970.

Fries, C.C. The Structure of English: An Introduction to the Construction of English Sentences. New York: Harcourt Brace & Co., 1952.

Gardin, J.C. 'SYNTOL" Rutgers Series on Systems for Intellectual Organization of Information. Ed. by Susan Artandi, Vol. II. Rutgers, New Brunswick: Thee State University, 1965.

Gardin, J.C. "Document Analysis and Linguistic Theory." Journal of Documentation 29(2); 1973.

Garfield, E. "Citation Indexes for Science." Science 122, 1955.

Garfield, E. "Science Citation Index: A New Dimension in Indexing." Science 144; 1965.

Grolier, Eric de. A Study of General Categories Applicable to Classification and Coding in Documentation. Paris: UNESCO, 1962.

Groupe d' Etudes Documentaires pour les Sciences Humaine of the Association. Marc Bloch Report I, Page I.

Hartmann, R.R.K. and Stork, F.C. Dictionary of Language and Linguistics. London: Applied Science Pub., 1972.

Hutchins, W.J. Language of Indexing and Classification. Herts, England: Peter Pereginus, 1975.

International Organization for Standard ISO 2788. Guidelines for the Establishment and Development for Monolingual Thesauri, 2nd Ed. Geneva: ISO, 1986.

International Organization for Standard, ISO 5964. Guidelines for the Establishment and the Development of Multilingual Thesauri. Geneva, ISO, 1985.

Jones, Spark Karen. Information Retrieval: A Draft for Encyclopaedia of Artificial Intelligence, 1990.

Kaiser, J.O. Systematic Indexing. London: Pitman, 1911, (The Card System Series), Vol. II.

Katz, J.J. and Postal, P.P. An Integrated Theory of Linguistic Description. Cambridge: The MIT Press, 1964.

Kyle, Barbara. "Towards a Classification for Social Science Literature." American Document, V.9; 1958.

Lancaster, F.W. Information Retrieval System: Characteristics, Testing and Evaluation. New York: John Wiley and Sons Inc., 1972.

Library Association Record. "The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval." Library Association Record, 47:7; 1955.

Luhn, H.P. "Keyword-in-Context for Technical Literature (KWIC Index)". American Doc. 11; 1960.

Luhn, H.P. (1952-64). "Selected Works." In H.P. Luhn: Pioneer of Information Sciences, ed. by C.K. Schultz. New York: Spartox, 1968.

Lyons, John. Introduction to Theoretical Linguistics. Cambridge: At the University Press, 1968.

Montegomery, C.A. "Linguistic and Information Science." Journal of American Society and Information Science, 23(3); 1972.

Mooers, Calvin M. "A Mathematical Theory of Language Symbols in Retrieval." In International Conference on Scientific Information, Vol. II. Washington: National Academy of Science, 1959.

Mooers, Calvin M. The Indexing Language of an Information Retrieval System. Papers Presented at an Institute Conducted by the Library School and the Centre for Continuation Studies, ed. by Wisley Simontor, Minneapolis Minn: The Centre, 1963.

Needham, C.D. Organizing Knowledge in Librariees: An Introduction to Information Retrieval, 2nd Rev. Ed. London: Adre Deutsch, 1971.

Neelamghan, A. and Maitra, R. "Non-Hierarchical Associative Relationships in Social Sciences: Identification & Typology." DRTC Seminar, V. 15, 1977.

Neelameghan, A. "Research on the Structure and Development of Universe of Subject." Library Science, V. 4, 1967.

Niel, S.D. Dilemmas in the Study of Information. New York: Greenwood Press, 1992.

Paillet, Jean-Pierre & Dugas, A. Approaches to Syntax. Amsterdam: John Benjamin Pub. Co., 1982.

Perry, J.W., Kent, A. and Berry, M.M. Machine Literature Searching. New York: Interscience Pubs., 1956.

Perry, J.W. and Kent, A. Documentation and Information Retrieval. Cleveland: Western Reserve University, 1957.

Perry, J.W. et al. Tools for Machine Literature Searching. New York: Interscience Pub., 1958.

Popper, Karl. Conjectures and Refutations. 4th Ed. London: Routledge and Kegan Paul, 1972.

Ranganathan, S.R. Colon Classification, 6th ed. Madras Library Association. London: Asia Publishing House, 1960.

Ranganathan, S.R. A Descriptive Account of Colon Classification. Bombay: Asia Publishing House, 1965.

Ranganathan, S.R. Prolegomena to Library Classification, 3rd Ed. Bangalore: Sarda Ranganathan Endowment for Library Science, 1967.

Richards, J. Longman Dictionary of Applied Linguistics. Essex, England: Longman, 1985.

Spark, Jones K. and Kay, M. Linguistic and Information Science. New York: Academic Press, 1973.

Soergel, D. Organizing Information, Principles of Data Base and Retrieval Systems. London: Academic Press, 1985.

Taube, M. Studies in Coordinate Indexing. 4 Vols. Washington: Documentation Incorporated, 1953-57.

Taube, M. "Notes on the Use, Roles and Links in Coordinate Indexing." American Documentation 12(2), 1961.

The UNESCO: IBE Education Thesaurus, 2nd Ed., Paris: The UNESCO Press, 1975.

Vickery, B.C. Faceted Classification: A Guide to Construction and Use of Special Schemes, London: Aslib, 1960.

Vickery, B.C. On Retrieval System Theory, 2nd Ed. London: Butterworths, 1965.

Vickery, B.C. Classification and Indexing in Science, 3rd Ed. London: Butterworths, 1975.

Vickery, B.C. "Knowledge Representation: A Brief Review." Journal of Documentation 42(3); 1986.