CHAPTER D

DEEP STRUCTURE OF SUBJECT INDEXING

LANGUAGES

• Components of SIL

• Structure of SIL statements
Derivation of meaning representations is a prime concern of the work going on in artificial intelligence and specifically automated language processing and is of practical and methodological interest to document retrieval. Chomsky, Fodor, Katz and other structural linguists and generative grammarians proposed the existence of 'universal linguistic forms'. (See Section B6). Chomsky (90) has pointed out that "real progress in linguistics consists in the discovery that certain features of given languages can be reduced to universal properties of language, and explained in terms of deeper aspects of linguistic form". Also he has observed that "more intriguing, to me at least, is the possibility that by studying language we may discover abstract principles that govern its structure and use - principles that are universal by biological necessity and not mere historical accident that derive from mental characteristics of the species" (91). It has also been observed by De Grolier (92) that fundamentally, the general semantic structure which is hypothesized for the Indo-European roots would be equally valid for other families of languages. Juret himself suggested that a parallel research would profitably be conducted on Chamito - Semitic languages. Roman Stupa's
investigations on the origin of language point to the fact that all human languages derive from a common foundation, which leaves the possibility that they partake uniform (or at least comparable and compatible) universal semantic structures — as it seems that they have common universal deep syntactic structures*. Such deep structure of sentences determine the semantic content while their surface structures determine the phonetic interpretation. In Bailey's theory, child and adult speech converge beautifully at the most crucial level — at the level of Deep Structure, where meaning is organised — and diverge elsewhere, at the level of sound (93). Such universals are being arrived at and used in various other fields dealing with information and information processing. In the area of data modelling, now the basic problem is to identify the world as a domain of objects with properties and relations (94). Other examples have been mentioned in paragraph B6.8.

With respect to documents, the aim of meaning representation is some degree of normalisation, which is of course what document description directly, or indirectly via requests is aiming at. In other words, all depend on
some sort of underlying semantic categorization and 'deep' syntactic relations \((95, 96)\).

D1 COMPONENTS OF SIL

As mentioned earlier in paragraph B7.1, the components of a SIL arc:

1. the vocabulary of the subject given in a natural language extracted as they occur and/or redefined for the purpose of subject analysis;

2. a set of relational data provided with the vocabulary forming the paradigmatic structure reflecting hierarchical relations, synonymous relations etc; and

3. a set of rules of syntax for formulating admissible subject headings, constituting the syntagmatic structure of the SIL.

D2 STRUCTURE OF SIL STATEMENTS

D21 Semantic Structure

Structure of a subject essentially refers to its species, parts and their interrelationships. The name of
a subject comprehends all its divisions and subdivisions.
In other words, the name of a subject summarizes indicative all its parts. What a name of a subject stands for, can be recognized on the basis of its denotation. The structure of a subject recognized on the basis of its denotation or comprehension is its "semantic structure".
For example, Mathematics as a subject comprehends Arithmetic, Algebra, Analysis, Geometry, Topology etc. Similarly, Agriculture as a subject comprehends Field Crops, Rice, Cultivation of Rice, Manuring of Rice, Harvesting of Rice etc. This structure is based on "Genus-Species", "Whole-parts", "Broader subject-narrower subject" or "Extension or range" of the subject — 'which has for its measure the number of existents or range' (67).

D22. Elementary Structure

In the name of a subject, the different substantive constituents (component elements) may belong to different categories depending on the semantic significance of the substantives. The structure recognized on the basis of the elementary categories to which the different substantives belong is the "Elementary Structure". This structure is artificially postulated. For example, according
to Kaiser's 'systematic indexing' (97, 98), the components of any name of subject are postulated to belong to 'Concrete' 'Place' and 'Process'. The categorization of components in a SIL serves a function quite similar to that performed by parts-of-speech or grammatical categorization of words in a natural language. (See paragraph 86.9). Words in a natural language such as English are viewed as belonging to categories such as noun, verb, adverb etc. The syntax of English grammar may then be defined with respect to these categories (99). In an analogous manner the syntax of expressions in a SIL may be defined with respect to the Elementary Categories.

D23 Syntactic Structure

The components in the name of a subject have relationships among them. The relationship is indicated by auxiliary words or function words or by their order. Changing the order of components result in changing the meaning they represent. For example, it has been shown in section C44 that "Destruction, Paint,
Bacteria" may mean "destruction of paint by bacteria" or "destruction of bacteria by paint". To preserve the meaning of the name of the subject without auxiliary words or function words, the order (sequence, syntax) of the different components are given as rules of syntax, with reference to the Elementary Categories. The structure of the name of a subject due to these rules of syntax gives rise to its "Syntactic Structure".

It is also of interest to note here, that the different components reflect the order, depth or intension (see section C3) of the subject they represent. This order, or depth or intension (the number of characteristics used in deriving the subject from its original universe (67)) is roughly proportional to the number of components used to represent it. In other words, as stated in section C3, representation of the name of a subject as per Elementary Categories according to a syntax is an order-preserving transformation, representing exactly the meaning of the subject and its order or intension or depth.
Evidently, the elementary structure and syntactic structure are closely related. The responsibility of preserving the meaning of the name of subject in the artificial subject indexing language rests mostly on them. The semantic structure depends mostly on the index terms, used to represent the components in the name of subjects, and their denotation in the natural language. Whereas, the elementary structure and syntactic structure are peculiar to the subject indexing language and are postulated to achieve the purpose of exhaustive and expeditious retrieval of information. The structure of a specific subject indexing language may be deemed to be a 'surface structure' of the 'deep structure' of subject indexing languages. For example, the structure of Kaiser's systematic indexing language has the elementary structure:

Concrete, Place, Process;

the syntactic structure being the same order (sequence). Ranganathan's subject indexing language has the elementary structure:

Basic Subject, Personality, Matter, Energy,
Space, Time, Common Isolate.
EXHIBIT - 3: DEEP STRUCTURE OF SUBJECT INDEXING LANGUAGE
D = Discipline
E = Entity
P = Property
A = Action
m = Modifier
The syntactic structure of Ranganathan postulates the categories Personality, Matter and Energy to repeat as Rounds after the Basic Subject, and the categories Space, Time and Common Isolate to occur at the end.

D24.1 By logically abstracting the structures of subject indexing languages of Cutter, Kaiser, Dewey, Ranganathan etc., a "Deep Structure of Subject Indexing Languages" has been arrived at (2). The Deep Structure of Subject Indexing Languages (DS of SIL) is presented schematically in Exhibit-3.

D24.2 The component ideas in the name of a subject according to the DS of SIL are deemed to fall in (belong to) any one of the Elementary Categories: Discipline, Entity, Property and Action, and a special component called 'Modifier'.

DISCIPLINE: An Elementary Category that includes conventional fields of study, or any aggregate of such fields or artificially created analogous fields. For example: Ocean Science, Physical Science, Physics, Biology, Leather Technology, Linguistics etc.
ENTITY: An Elementary Category that includes manifestations having perceptual correlates, or only conceptual existence, as contrasted with their properties and actions performed by them or on them. For example: Light, Chlorine, Plant, Animal, Gold, Energy etc.

PROPERTY: An Elementary Category that includes manifestations denoting the concept of "attribute" - qualitative or quantitative. For example: Capacity, Efficiency, Resistivity, Flexibility, Utility, Valency, Specific gravity etc.

ACTION: An Elementary Category that includes manifestations denoting the concept of "doing". Action may manifest as Self Action or as External Action. For example: Function, Migration etc., are Self Actions, Selection, Evaluation, etc. are External Actions.

NOTE: The term 'manifestation' has been used to denote an 'instance' or 'occurrence' of the respective Elementary
Categories. In other words, it has been used to refer to a specific idea falling in any one of the Elementary Categories. The names of the Elementary Categories are also used to denote their respective manifestations in this thesis.

**MODIFIER**: In relation to the manifestation of any one of the Elementary Categories, 'modifier' refers to an idea used or intended to be used to qualify the manifestation without disturbing its conceptual wholeness. For example: 'Infections' in 'Infectious Disease', 'X-ray' in 'X-ray Treatment'. A modifier can modify a manifestation of any one of the Elementary Categories, as well as a combination of two or more manifestations of two or more Elementary Categories. Modifiers generally create Species/Types. Modifiers can be Common Modifiers like Form, Time, Environment and Place, and Special Modifiers which can be Discipline based, or Property based, or Action based.
The directions shown by arrows in the schematic representation of the DS of SIL in Exhibit - 3, indicate that any manifestation of any one of the Elementary Categories may be related to any of another category, provided the product of their relation is meaningful.

D25 Rules of Syntax Associated with DS of SIL

The basic rule of syntax associated with the Deep Structure of SIL is that, Discipline should be followed by Entity (both modified or unmodified) appropriately interpolated or extrapolated wherever warranted by Property and/or Action (both modified or unmodified).

D25.1 A manifestation of Property follows immediately the manifestation in relation to which it is a Property; a manifestation of Action follows immediately the manifestation in relation to which it is an Action — are the other associated syntax rules. It has also been stated that, Property and Action can have other Property and/or Action directly related to it. In other words, the rule of syntax relating to Property and Action of another Property or Action is that, their positions are always after the Property or Action to which they are related.
The rules of syntax relating Species/Part and Modifiers are as follows: A Species/Part follows immediately the manifestation in relation to which it is a Species/Part. A modifier follows immediately the manifestation in relation to which it is a Modifier. This rule also applies to cases where there are more than one modifier to the same manifestation. If more than one sequence of modifiers to the same modifyee is equally valid (in terms of its representation in the natural language) according to the above rule, the choice of any one sequence is acceptable.

The above rules of syntax relating to the sequence of Elementary Categories form a significant contribution to the theory of subject indexing languages. As per the above rules, only the positions of Discipline and Entity are fixed. The positions of both Property and Action are not fixed and they are given a 'floating' position, in the sense that they go adjacently next to the manifestation to which they are respectively Property and Action. Moreover, Property and Action can go with
another Property or Action also. That is, there can be an Action on/of another Action without any other manifestation coming in between. This has removed the rigidity due to the concept of 'Rounds' in Ranganathan's Colon Classification and its associated Chain Indexing Language.  

D25.4 Another significant rule of syntax associated with the DS of SIL is as follows: In the context of constructing subject headings according to the DS of SIL, when the purpose is to group together in the subject index all or major portion of information pertaining to a manifestation or manifestations of a particular Elementary Category or Combination of Elementary Categories, the manifestation or Elementary Category or combination of Elementary Categories, as the case may be, can be taken to be the first Context Specifying Category instead of Discipline, and is called the Base. When the purpose is to further group together in the subject index all or major portion of information pertaining to manifestations of one or more Elementary categories, the Category or Categories concerned is/are taken to be the second category following the Base, called
Core. In the case of Kaiser's language the Elementary Category Entity is the Base. In Dewey's language (DDC along with its relative index), for example, Medicine which is a Discipline is taken as the Base. But the Core consists of Anatomy which is a Property, Physiology which is a Self-Action and Disease which is again a Property.

D25.5 The above rule of syntax associated with the D3 of SIL provides the facility of structuring the subject headings according to specific purposes, and to generate specific purpose oriented groupings in the subject index. Stated simply, this rule allows an indexer to decide which of the Elementary Categories should become the first context specifying category and second context specifying category respectively. Once this is decided the other rules of syntax guide the formulation of subject headings using the categories of the Deep Structure.

**NOTE:** The synonyms, quasi-synonyms and antonyms of the terms used in the subject headings are to be controlled by using Cross References. This is done by "standardization" discussed in section S3 in the
next chapter. Any vocabulary control tool like thesaurus could be used for this purpose, but a new type of such a tool called 'classaurus' would be the most appropriate, whose design and use are discussed later in chapter H.

The design, development and operation of a novel computer-based subject indexing system based on the DS of SID, named 'Deep Structure Indexing System', which is capable of generating different types of subject indexes from a single subject heading formulated according to DS of SID is described in the next two chapters.