

Chapter 2

The nature of documentary languages

- 2.1 Information systems are concerned with the communication of information about documents to the potential users (readers) of those documents. The means of communication are the documentary languages (DLs) of the information systems. At the most abstract level therefore, the study of documentary languages is part of the general study of communication and sign systems, or 'language' in its most general sense.
- 2.2 The serious study of sign-systems began, it is generally agreed, at the end of the nineteenth century and the beginning of the twentieth with the work of Charles Sanders Peirce and Ferdinand de Saussure. Known as semiotics or semiology – the latter particularly in France – the general science of signs has as its goal a general theory of signs in all their forms and manifestations, whether in animals or in man, whether normal or pathological, whether linguistic or nonlinguistic, whether personal or social (Nauta, 1972, p.26).
 - 2.2.1 The central concept of semiotics is the process of 'semiosis', defined as a five-place relation $S(s, i, e, d, c)$ and described by Morris (1938) as a 'mediated taking-account-of'. The mediators in the process are sign-vehicles ('s'); the takings-account-of are the effects ('e') of the process; the senders or receivers of the process are interpreters ('i'); what is taken account of are designata ('d'); and the external factors influencing the process are the context ('c'). The sign-vehicle is the particular object or event, such as a sound or a mark, that functions as a *sign* for an organism (the interpreter). In its function as a sign, the sign-vehicle generally refers to something, the denotation or referent ('r'), outside the semiotic process proper. The effect ('e') of a sign-vehicle is to dispose the receiver ('i') to react in a certain way to the designation (signification or sense) of the sign ('d'), i.e. to identify the referent intended.
 - 2.2.2 Morris illustrated semiosis by an example from the 'language of bees'. The specific dance functions as a sign ('s'), the bees affected by the dance are the interpreters ('i'), the disposition to look for food in a certain direction and at a certain distance is the effect ('e') of the sign; 'food' (the special kind of object that bears food) together with the directives for the search for food is the designatum (signification) ('d') of the sign, and the actual food-objects sought are part of the referent ('r') of the sign; and lastly, the position of the hive is part of the context ('c') of the sign.
 - 2.2.3 In natural language (NL) communications the sign ('s') is the sequence of physical sounds or written marks; the interpreters ('i') are the speakers (writers) and hearers (readers); the designata ('d') are the relationships between the physical forms of signs and the physical objects they refer to, their referents ('r'); the effects ('e') of signs are the changes evoked by designata in the dispositions of interpreters; and the contexts ('c') are the textual and situational environments in which particular acts of communication take place.

2.2.4 Preliminarily, we may characterise the semiosis of an indexing or classification system as follows. The sign ('s') is the sequence of physical forms of the documentary language (DL): the interpreters ('i') are the indexers and index-users (enquirers); the designata ('d') are the relationships between the physical forms of the DL signs and their referents ('r'), the subject content of documents; * the effects ('e') are the reactions of enquirers (e.g. to look for the document or not according to their judgement of its relevance); and the contexts ('c') are the physical arrangements and general environment of the index and of the information centre etc.

2.2.5 When studying a semiotic process, we may adopt different approaches according to the needs of the moment. First, we may study the formal aspects of signs where the chief emphasis is upon the form of 's' and upon the ways in which complex sign-vehicles may be constructed out of elementary ones. Secondly, we may abstract from the process the elements 's' and 'd' and concentrate on the semantic aspects: the relationships between a sign ('s'), its designatum ('d') and its referent ('r'), if any; the ways in which designata may be related to each other; and how they may combine in giving 'meaning' to complex sign combinations. The third aspect is the functional or pragmatic aspect. It is the study of the effect ('e') of signs upon interpreters ('i') and the contexts ('c') in which signs are used. Studying the use and effect of the information conveyed by signs presupposes that signs have meaning, i.e. a study of the pragmatic aspect of semiosis presupposes a study of the semantic aspect, which in turn of course presupposes a study of the formal aspect. A description of a semiotic system at the pragmatic level concentrates essentially upon the elements 'i', 'e' and 'c' but it cannot ignore the elements 's' and 'd'. Equally, a description at the semantic level concentrates on the relationships between 's' and 'd' but it cannot entirely omit the formal aspects of 's'.

2.3 In describing natural language, linguists often draw a fundamental distinction – first introduced by de Saussure ([1916], 1964) – between *langue* and *parole*. The set of utterances produced by speakers of a language are instances of *parole*, which the linguist takes as evidence for the construction of the underlying system potentially available to all speakers, which is *langue*.

A similar distinction has been drawn by Chomsky (1957, 1965) between linguistic *competence* and linguistic *performance*, between what every speaker is capable of, given ideal and complete knowledge of the language, and what speakers actually produce. The dividing line comes effectively between the formal and semantic levels on one hand and the pragmatic level on the other.

2.4 Within the language-system (*langue*) itself, linguists generally identify for natural language a number of levels or strata, not just only a formal level and a semantic level.

2.4.1 First, there is the distinction made between phonetics and phonemics (or phonology). Although the *t* in *but* is quite different acoustically from the *t* in *teeth*, no English speaker would fail to regard them both as instances of the 'same' sound *t*. Nor would he fail to distinguish between the *t* in *better* and the *d* in *bedder*, even though in many dialects they are acoustically almost identical. Such facts are accounted for by saying first that the several phonetically distinct *t*'s are all variants of a single phoneme /*t*/ (or alternatively that one phoneme /*t*/ is realised variously by different phonetic forms according to the context in which it occurs); and secondly that one variant of /*t*/ is phonetically similar to one variant of /*d*/. Whereas phonetics is the science of describing the actual sounds of speech, phonemics is the description of what phonetic features are important for speakers of the language in distinguishing and discriminating between phonemes within a particular language-system. Such distinctive phonetic features are, for example, vocalicity and tongue-position (dental, labial, palatal, velar etc).

* Note that the referents of DL signs are not the documents themselves but their contents (4.2.3 below).

One phoneme is discriminated from another by at least one distinctive feature, e.g. the two dental plosives /*t*/ and /*d*/ differ in vocalicity – /*d*/ is 'voiced', /*t*/ is not. In other words, phonemes are 'bundles' of distinctive phonetic features.

2.4.2 The stratum above the phonemic is generally known as the lexical or lexemic. A lexical unit, or lexeme, is an element composed of a string of phonemes having significance, i.e. it is a sign ('s') having a designatum ('d'). Lexemes combine to form phrases, clauses and sentences according to the language rules traditionally known as syntax. It is evident that such complexes as phrases, clauses, sentences and indeed whole texts, may be considered at the phonemic level as various combinations of phonemes and at the lexemic level as various combinations of lexemes. Each stratum therefore has its own basic unit (phoneme, lexeme) which can be combined to form units of higher 'ranks' (Lyons, 1968).

2.4.3 We must be careful to distinguish lexemes from 'words'; or rather, we must be clear in what sense a word may be said to be a lexeme. With Lyons (1968, p.196ff), we can identify three uses of 'word': the phonological word, the orthographical word and the grammatical word. For example, the phonological word /*sæt*/ and the corresponding orthographic word *sat* represent a particular grammatical word, which is traditionally referred to as 'the past tense of *sit*'; whereas the phonological word /*kʌt*/ and the corresponding orthographic word *cut* represent three different grammatical words, 'the present tense of *cut*', 'the past tense of *cut*' and 'the past participle of *cut*'. Generally, there is a one-to-one correspondence between phonological and orthographic words, but there are some instances of many-one and one-many correspondences, e.g. /*mi:t*/ : *meet* or *meet*; and, /*ri:d*/ or /*red*/ : *read* (i.e. 'the present tense of *read*' or 'the past tense of *read*').

2.4.4 A lexeme is a more abstract notion related to grammatical words. We said that the orthographic word *cut* corresponds to three grammatical words, namely, the various tenses of 'cut'. Alternatively, we can say that these three grammatical words are the various realisations (in particular grammatical contexts) of the single lexeme CUT. More commonly, the various grammatical realisations of a lexeme have different morphological forms, e.g. the lexeme SING appears in the variants (grammatical words) *sing*, *sings*, *sang*, *sung*, *singing*. The lexeme itself is therefore an abstraction devoid of any grammatical function. Its formal realisations are the result of the addition of phonemes or groups of phonemes to the lexemic 'root' (the phonemic unit constituting the lexeme). Such phoneme groups having grammatical significance are often called (by French linguists particularly) 'morphemes'.

2.4.5 For many linguists, however, morphemes are the units of a stratum intermediate between the phonemic and the lexemic. They are then defined as the smallest combinations of phonemes having significance within the language system. For example, we may say that the (phonological/orthographic) word *unacceptable* is composed of three morphemes *un*, *accept* and *able*, three phonemic sequences, each of which has a particular distribution on the lexemic stratum, i.e. may occur with other morphemes in the constitution of different words (*undetectable*, *acceptance*, *respectable*, *untrue*, etc.). Some morphemes are 'bound', they occur only in combination with other morphemes, e.g. *un-*, *-ed*, *-ly*. Others are 'free', they may occur alone – as realisations of a single lexeme (e.g. *accept*, *detect*, *respect*). The description of morphemes and how they combine to form words and lexemes is known traditionally as morphology.

2.4.6 Morphemic analysis is not free of theoretical controversy. The chief dispute concerns the criterion of significance. Some linguists would recognise a morpheme *-ceive* occurring in words such as *deceive*, *receive*, *conceive*, etc., but it is doubtful whether the morpheme has a designatum, in the sense that lexemes do. We may say that morphemes have meaning insofar as they can be identified with parts of the designata of the lexemes which they compose, e.g. *re-* has the general sense of return or repetition. In the case of the morphemes added to lexemic roots the identification of a morphemic

meaning is much clearer. We might say that *-s* is the morphemic expression of 'plural' in English, *-ly* is the morphemic expression of 'adverbiality'. Where morphemes coincide with lexemic roots then it is obvious that they do have signification — not as morphemes, but as lexemes. For many morphemes however, no meaning can be identified — our example of *-ceive* seems to be one. The true status of morphemes is thus very much in dispute, and to pursue it further would take us well beyond the primary needs of this study.

- 2.4.7 The next stratum, above the lexemic, is concerned with the designata of utterances and their interrelationships. It is now frequently called the sememic stratum. Its basic unit, the sememe, may be defined as the signification of a lexeme (or a morphemic unit such as the plural *-s*), i.e. it represents the relationship between a sign ('s') and its referent by virtue of which the sign may partake in semiosis. It is a unit exclusively within the language-system — considerations at the pragmatic level are ignored. That is to say, the 'meanings' a particular lexeme (word) may have for a particular speaker, its connotations, are not the province of the sememe. A sememe is thus defined as the signification (designatum) of a lexeme for the standard or 'ideal' interpreter. (We shall elaborate on these distinctions and on related questions in ch. 4 below.)

Just as lexemes combine to form phrases, sentences and texts, so do their corresponding sememes form the semantic representations underlying phrases, sentences and texts. Similarly, just as phonemes are analysable in terms of distinctive features, so may sememes be analysed in terms of primitive semantic components.

- 2.4.8 In a sense, primitive semantic components provide the link between the sememic stratum and the extra-linguistic world of referent-objects and events. They are the linguistic counterparts (i.e. internal to the language-system) of those conceptual or perceptual characteristics, properties, attributes, etc. of the objects, ideas, events, feelings, etc. about which one communicates through language. The relationship between the sememic stratum and the non-linguistic world of referents is similar to that between the phonemic stratum and the non-linguistic world of physical sounds or written marks. Just as when listening to a stream of vocal sounds, we pick out phonemic units on the basis of discriminable features (e.g. vocality, nasality, dentality, etc.) so, when selecting words to express ourselves, we pick lexemes on the basis of the semantic components in their associated sememes which correspond to the features discerned (discriminated) in the referents of our intended utterance.

- 2.4.9 On each of these three (or four) strata, elements are related to each other on two basic axes. First, as we have seen, elements may be combined to form units of higher rank: phonemes may be linearly concatenated in syllables or morphemes, lexemes may be combined by syntactic rules into phrases and sentences, and sememes may be joined together in their underlying semantic representations. The combination of elements takes place on the *syntagmatic* axis, and the complexes that are formed are called 'syntagms'. As opposed to this, elements are related to each other according to their potential mutual substitutability in syntagms. This, the *paradigmatic* organisation of elements, is determined by the sharing of common features. For example, the substitution of /p/ for the phoneme /b/ in *bet* is made possible by both phonemes sharing the same distinctive feature (labiality) and differing only in that /p/ is voiceless while /b/ is voiced. Similarly, the substitution of one lexeme for another may be determined either by the sharing of common morphemes (particularly those realising grammatical functions, e.g. plurality, past tense, etc.) or by the sharing of common semantic components in their respective sememes. From one point of view, we may regard the paradigmatic axis as analytical and the syntagmatic axis as synthetic. The paradigmatic relationships are 'fixed' by the structure of the language-system, whereas the syntagmatic ones are 'selected' during the utterance of particular messages — the former are *a priori* while the latter are *a posteriori*. The clearest example is found on the lexemic stratum in the opposition between vocabulary and syntax, where vocabulary represents the set of lexemes available to speakers and syntax represents the rules governing their combination into sentences.

- 2.5 Distinctions of a parallel nature are to be found in documentary languages (DLs). It is now a well established convention to talk of vocabularies of descriptors and of the syntactic rules for combining descriptors to represent document content in index files. Descriptors correspond, therefore, to the lexemes of natural language (NL). They have both form and meaning. At the formal level they can be described in terms of syntagms of notational symbols (numbers, letters, punctuation marks) — parallel to the description of NL lexemes as combinations of phonemes. And at the semantic level, corresponding to the sememes of NL lexemes are the designata of descriptors, which like NL sememes may be analysable as syntagms of (primitive) semantic components.

- 2.5.1 These similarities will be elaborated upon and refined in succeeding sections, but at this point we must bring out the important differences between DLs and NLs. At the formal level, the most obvious one is that while in NLs the written forms are secondary to the vocal forms and are usually derived from them (the exceptions being the hieroglyphic and ideogrammatic scripts, e.g. of Egyptian and Chinese, respectively), in DLs it is the written forms that are basic. If they can be spoken it is only derivatively, i.e. by analogy with relationships in NL between written and spoken forms (3.2.2 below).

- 2.5.2 At the semantic level, DLs differ from NLs in standardising or simplifying the relationship between the lexemic and sememic strata. The existence of synonyms and homonyms in NL shows that there is no one-to-one correspondence between lexemes and sememes. But DLs are designed, as we shall see, to eliminate, or at least greatly reduce, homonymity and synonymy. In other words, DLs attempt to unify the lexemic and sememic strata. Consequently, it is legitimate in the case of DLs to refer simply to a single semantic level of description, as we do in chapters 4 and 5.

- 2.5.3 A final basic distinction, not so far touched upon, is that whereas NLs can function as their own metalanguage, DLs cannot. A metalanguage is defined as the language used to talk about another language, the object-language. The English language, for example, can be used as a metalanguage to talk about mathematical symbols, the object-language of mathematics, e.g. *the sign '+' means addition*. But English, like other NLs, can also be its own metalanguage, e.g. *the word "boy" is spelled with three letters*, *"Boy" means a young male human being*, etc. Like other artificial languages, DLs cannot function as their own metalanguage. Another language must be used, either (as in this book) a natural language or, more restrictedly, another artificial language, e.g. symbolic logic or the symbolism of a linguistic theory (as we also do here on occasion).*

- 2.6 Documentary languages are clearly different from natural languages, and seem to fall in the category of artificial languages. The essential and defining characteristic of an artificial language is that, in contrast to NL, it has been designed for a specific purpose or range of functions. As a corollary to this, the ways in which it is to be used are laid down — the rules governing its use are *de jure* as opposed to those of NL which are *de facto*. "*De facto* rules are rules that state something that happens, irrespective of whether it ought to or not; ... *de jure* rules ... state something which may or ought to happen" (Nauta 1972 p.49). The grammatical rules of English are descriptions of how most people produce or interpret English utterances; they describe the language-system as it is known by the ideal speaker-hearer (2.3 above), not as it 'ought' to be. It is certainly true that in the past many grammarians have 'prescribed' correct usage, but even then, they were describing what some speakers of the language did at the time or in some 'more perfect' past. The apparently *de jure* statements of authorities such as the standard dictionaries and books on style and usage (e.g. Fowler, 1926 and Gowers, 1948), are, in general, to be regarded in this light. The fact that speakers consult such authorities does not mean that the NL they speak is governed by *de jure* rules.

* We can argue that DLs are metalanguages insofar as they are used to talk about the (NL) texts of documents (4.2.4 below). But even if this is true, it is still not the case that DLs can be their own metalanguage, i.e. 'a metalanguage of a metalanguage'.

2.6.1 By contrast, the use of artificial languages such as symbolic logic, one of the prototypes of an artificial language, is completely determined *de jure*. In the formulation and interpretation of its formulae users must follow the rules prescribed; deviation results automatically in misuse of the language. The breaking of the *de facto* rules of NL has, by contrast, no such drastic consequences: the rules are persistently broken with rarely much loss of comprehension — as witness much of poetry.

2.7 Artificial languages are designed to fulfil a particular function within the possible range of semiotic functions as effectively as possible. Natural languages do not have specific functions, they function in many different contexts and for many different purposes. It is this very heterogeneity in their functional (pragmatic) aspect that makes them less efficient, or even inefficient, when used for more limited functions. In a very real sense, it is true to say that the characteristic structures of artificial languages are determined largely by the functions they are designed to perform.

2.7.1 We may summarise the principal functions of NL as follows*

- (i) the communication of factual information; the descriptive or informative function. This function is frequently but erroneously assumed to be the only purpose of linguistic communication, and it is the source of many misconceptions about the nature of 'meaning' (4.13). Although it cannot be denied that most NL utterances have some factual or 'ideational' content (Halliday 1973), it is very rare to find utterances lacking one or more of the following 'interpersonal' functions:†
- (ii) the promotion or incitement of specific behaviour in recipients; the incitive, conative, imperative or manipulative function(s).
- (iii) the indication of value-judgements, by the speakers, on events, actions or purposes; the valuative function.
- (iv) the direction or guidance of recipients in the performance of an action; the prescriptive function.
- (v) the expression of feelings, experiences, moods, needs etc., of the speaker; the expressive function.
- (vi) the evocation of certain states of mind in the recipient; the evocative or aesthetic function.
- (vii) the making of social contacts, e.g. in greetings, 'small talk'; the phatic function.
- Lastly, one function already mentioned and not involving interpersonal relations:
- (viii) the discussion of the language itself; the metalingual function.

2.7.2 Artificial languages are generally limited to only one or two of these functions. For example, the language of a musical score and the code of traffic symbols are both prescriptive languages. Computer programming languages are essentially incitive and prescriptive; a statement in FORTRAN is an instruction (command) that the computer perform a certain operation. The great majority of artificial languages, however, are restricted to the descriptive function, the communication of information, and very often they are limited within this function to specific facets. The languages of symbolic logic concentrate on the syntagmatic axis (the combination of lexical items) as in the predicate and propositional calculi. Other languages concentrate on particular subject fields, e.g. the languages of chemical formulae, or, like DLs, on a particular channel of communication, i.e. documents.

* Based largely on Bühler (1934) and Jakobson (1960)..

† Since interpersonal functions necessarily involve the participants of a discourse — the interpreters (i) — their description inevitably takes place at the pragmatic level. By contrast, the ideational function is almost exclusively confined to the semantic level (cf. Halliday 1970a, 1973).

2.7.3 Whatever their particular function, artificial languages are designed to be more 'efficient' than natural language. They are concerned primarily to reduce, or even eliminate, the redundancy and ambiguity of NL; the inevitable consequences of the ways in which NLs are transmitted from one generation to another, of its 'noisy' vocal-oral communication channel, of speakers' and recipients' less than perfect decoding capacities, of its almost infinite variety of functions and uses, and so forth.

Artificial languages seek to 'normalise' NL semantics. Where a NL may have two or more expressions for the same object or concept, artificial languages aim to have only one expression. One statement of a computer programming language, one musical notation, one chemical formula might be 'translatable' by a variety of different NL forms. The most obvious way to achieve such normalisation is to create special symbols for the purpose, as in music and chemistry. An alternative is to develop a restricted form of NL, i.e. to use the forms of NL in a standardised, controlled manner, eliminating ambiguity and redundancy.

2.7.4 Both these methods are found in documentary languages (DLs). The use of a special notation gives us the classificatory languages (CLs) such as the Dewey Decimal Classification, the Universal Decimal Classification (UDC), the Library of Congress (LC) classification, etc. The use of a 'standardised' version of NL gives us the indexing languages (ILs), the various kinds of alphabetical subject indexes, coordinate indexes, and their auxiliary tools, the thesauri.*

2.8 The normalisation or standardisation of NL in one way or another is a common feature of artificial languages in general. Where DLs differ from other artificial languages is in the particular functions they have to perform, and in the influences that functional requirements, or, in more general terms, pragmatic factors, have upon their structures as language-systems.

2.8.1 First, there is the primary function of DLs as channels of communication between documents and potential readers. In principle any 'subject' of a document (however we may define 'subject', see ch.7 below) and any 'subject' of a request to the system, must be capable of expression in the DL. Except in the most restricted and specialised environments, the range of discourse with which a DL must be able to deal is considerable. Generally speaking, it is much greater than that of other artificial languages (with the exception of the international auxiliary languages)†. Consequently, whereas the users of chemical and musical notations are expected to learn to communicate directly in these languages, nobody expects the users of index files to learn the DL. A chemist can be expected to know the chemical symbols and to be able to interpret chemical formulae without constant reference to a key (a dictionary giving explanations in NL) because the referential extension of the language is quite narrow. But with DLs the range is too great and dictionaries must be provided for access from NL. In the case of CLs, such access tools take the form of 'bilingual dictionaries' giving the equivalents of NL expressions in the form of CL descriptors. In the case of ILs, they take the form

* In many publications the terms 'indexing languages' or 'index languages' are used in a much wider sense, to refer to all languages and codes found in information systems. While there are convenient terms for languages using classification symbols — i.e. classification schemes or (as here) classificatory languages — there is no suitable term for languages using NL forms other than 'index(ing) language' itself. To avoid confusion we have decided to use indexing language (IL) in a more restricted sense than normal, and to employ the (less usual) term 'documentary language' (DL) to embrace all types of language found in information systems.

† The international auxiliary languages such as Esperanto, Ido, Novial etc. differ from other artificial languages, including DLs, in a number of respects. Much greater emphasis is placed upon their phonemic features, there is greater allowance for redundancy, a certain degree of polysemy and ambiguity is permitted and they may, like NLs, be their own metalanguages. In brief an effort is made to give them the functional diversity of NL while maintaining a higher degree of normalisation.

of 'thesauri' showing when a NL form has been adopted as an IL descriptor and, if not, showing how it is expressed in the IL. Other functions may be incorporated (6.2 below) but the principal one for all such 'access tools' is to provide translations from NL into DL.

- 2.8.2 In the most general terms, the purpose of an information system is to supply enquirers with the documents that should be of interest to them. It is generally agreed (Coates 1960, Foskett 1971 and Vickery 1971) that there are two types of request that enquirers ask of index files: (i) they want to retrieve the documents in the system that are relevant to a specific topic of interest; and (ii) they want to have (a selection of) documents lying within, or embracing, a particular range of related or associated topics, whether constituting a recognised field of study or not. We may follow Vickery (1971) in referring to these two types of search request as 'specific reference' and 'generic survey', respectively.
- 2.8.3 For the efficient answering of 'specific reference' search requests the initial and fundamental requirement is that the DL can express precisely the subject contents of documents and the search requests themselves. Even in a restricted subject field, few DLs can supply specific individual descriptors to cover all contingencies. The majority of DLs must have the capacity to form descriptor syntagms for the expression of a (complex) subject – in other words they must have syntagmatic organisations of one kind or another. Basically, there are two methods; either the subject description of a document is formulated in terms of descriptors that are entered separately in an index file, or it is formulated in terms of a syntagm of descriptors – descriptor phrase (Chapter 3ff) – that is entered as a whole in the index file. In the first case, search requests specify a set of descriptors that are to be used in the coordination of entries in the index file; a *postcoordinate* system. In the second case, search requests must be formulated as syntagms (descriptor phrases) and matched as a unit against entries in the index file; a *precoordinate* system. In practice, however, the two approaches are not exclusive, since most precoordinate systems allow the coordination of descriptor phrases during searches (8.5.1; 8.6.1).
- 2.8.4 The syntagmatic organisation of a DL can therefore be a facet of both its indexing function and its searching function, as it clearly is in precoordinate systems, or it may be operative only in search processes, as in 'pure' postcoordinate systems. In the latter case it may be legitimate to talk of a 'search language' (Vickery 1971), 'searching component' or 'searching sublanguage', since the form of the DL in its searching functions is quite distinct from its form in the indexing functions, i.e. by the presence and absence, respectively, of a combinatorial (syntagmatic) facility (see further in ch.8).
- 2.8.5 Turning now to the second type of search, the 'generic survey', the requirement is that enquirers be able to consult documents on subjects related in some way to the subject stated in the search request, i.e. subjects broader (more generic) in scope, subjects narrower (more specific) in scope, or subjects in associated fields. For this to be possible, the descriptors of the DL need to be themselves related to each other generically, specifically and associatively, in an explicitly formulated paradigmatic structure. Examples are the classification hierarchies of CLs and the thesauri of IL descriptors.
- 2.8.6 The paradigmatic structure of a DL can however have a number of other additional functions, e.g. the guidance of indexers in the selection of descriptors for document descriptions and the similar guidance of enquirers in the selection of descriptors for search requests. In the case of CLs, there is normally a further important function – and one for which CLs have often been primarily devised in the first instance – namely, the physical ordering of documents on the shelves of the information centre (library, documentation service, etc.). This function has particular relevance to the formal characteristics of CLs (3.2.2).

- 2.8.7 The same is true for a further feature of DLs that is not essential to the definition of DLs as such, but is nevertheless a desirable property. This is its flexibility in the face of changes in the patterns of knowledge, of subject fields and of the ways in which they are written about in documents, and its accommodation of new perspectives and new attitudes towards traditional topics – in other words the 'hospitality' of a DL, reflected largely in its notational flexibility.
- 2.8.8 Pragmatic considerations play a large part in the design of any DL. Most important is that a DL should impose an organisation of documents and their contents that is generally acceptable to the great majority of the users for most of the time. The structure of a DL should, as far as possible, take account of the environment ('context') of the information system in which it is a crucial component, the types of users served ('interpreters') and their interests, the kinds of documents processed in the system and the range of subjects covered ('referents'). These factors are most easily identified in the case of DLs designed for specialised information centres serving a well defined clientele, but they are present also in the case of general DLs covering the whole range of knowledge and its documentation and serving users of all types and all levels of interest.
- 2.8.9 It is clear that the characteristic structural properties of DLs are determined largely by their particular functional requirements. As in all languages, whether natural or artificial, linguistic structure and linguistic function are interdependent. We cannot describe how a language works without describing its structure, and we cannot fully explain its structures without mentioning the functions it has to perform. To understand the linguistic processes in the indexing and retrieval of documents, we must also examine the linguistic structure of the DLs involved, and the logical sequence of such an examination is, as we have seen (2.2.5), to proceed from the formal level of linguistic description to the semantic and then to the pragmatic. At this point we can then return to a fuller and deeper description of the characteristic DL functions of indexing and searching.