# BIBLIOGRAPHIC CLASSIFICATION THEORY AND TEXT LINGUISTICS: ABOUTNESS ANALYSIS, INTERTEXTUALITY AND THE COGNITIVE ACT OF CLASSIFYING DOCUMENTS

CLARE BEGHTOL

*220 Heath St. E., Toronto, Ontario, M4T 1S9, Canada*

A strong definition of aboutness and a theory of its role in information retrieval systems have not been developed. Such a definition and theory may be extracted from the work of T. A. van Dijk. This paper discusses some of the implications of van Dijk's work for bibliographic classification theory. Two kinds of intertextuality are identified: that between documents classified in the same class of the same classification system; and that between the classification system as a text in its own right and the documents that are classified by it. Consideration of the two kinds of intertextuality leads to an investigation of the linguistic/cognitive processes that have been called the 'translation' of a document topic into a classificatory language. A descriptive model of the cognitive process of classifying documents is presented. The general design of an empirical study to test this model is suggested, and some problems of implementing such a study are briefly identified. It is concluded that further investigation of the relationships between text linguistics and classification theory and practice might reveal other fruitful intersections between the two fields.

## 1. ABOUTNESS

IN 1977 the Aslib Co-ordinate Indexing Group (CIG) held a colloquium at which six speakers presented their current views on aboutness. MacCafferty's report on the meeting concluded that 'we ought to establish some form of framework upon which *any* particular point of view on aboutness can be placed.'[1] Fairthorne, one of the speakers at the colloquium, had previously suggested such a framework by distinguishing between extensional and intensional aboutness in documents.[2] Extensional aboutness, in Fairthorne's terms, is the inherent subject of the document; intensional aboutness is the reason or purpose for which it has been acquired by a library or requested by a user. Others, too, have noted that any document has more than one subject, depending upon who wants it, why, and what for. Hutchins, for example, cautioned that 'we should never talk of *the* subject of a document' because the subject of a single document varies along a number of dimensions.[3] The Classification Research Group (CRG) faced this dilemma when, having devised an experiment to see if they could achieve consistency in classifying within a given system, members of the Group were unable to agree on the subjects of the documents. Eventually, the chairman unilaterally decided what the documents were about, the experiments were conducted and agreements were reached on the best classifications for the documents.[4]

The two kinds of document aboutness that Fairthorne identified can be distinguished more sharply by constrasting 'aboutness' (extensional aboutness) with 'meaning' (intensional aboutness). Boyce[5] made a similar distinction using

the terms 'topicality' (aboutness) and 'informativeness' (meaning). Boyce's terms seem to suggest that successful searches automatically provide new information when they may, in fact, simply confirm or fail to confirm information previously known to the enquirer. Van Dijk[6] distinguished between aboutness and meaning in a slightly different sense with the terms 'normal relevance assignment' and 'differential relevance assignment' for the special case in which one reader is reading a particular text. Van Dijk uses 'relevance' to mean the relationships perceived by readers between the various structural and semantic elements of a particular text while it is being read. Since this usage differs from the way 'relevance' is conventionally used in discussions of information retrieval systems, it seems better to avoid this possible source of confusion.

Whatever terms are chosen, a distinction between 'aboutness' and 'meaning', as the terms are used here, seems justifiable on the assumption that a document has an intrinsic subject, an 'aboutness', that is at least to some extent independent of the temporary usage to which an individual might put one or more of its meanings. This assumption is somewhat controversial. Stimulating debates upon its validity for literary texts can be followed in Valdés and Miller.[7] For the present purposes, however, we may take the general position that texts of all kinds have a relatively permanent aboutness, but a variable number of meaning(s). There is, of course, a strong relationship between a document's aboutness and its potential meanings for individuals, so the distinction should not be taken as a rigid one; but it clarifies the point that a document may have only one aboutness, but an unlimited number of meanings, differing according to the exact use a particular person may find for the document's aboutness at a certain time. Indeed, the same document can have different meanings for the same reader at different times, but the document, itself unchanging, is assumed to possess a fundamental aboutness.

A recognition of the relatively permanent quality of aboutness in documents is one of the assumptions upon which bibliographic classification systems have traditionally been based. Classificationists have endeavoured to create classification systems conceptually and notationally hospitable to any aboutness a document might present, but it has not been suggested that the inherent aboutness of the document changes when a particular meaning is attached to it or a particular use made of it by a reader. As Robertson pointed out, the purpose of retrieval systems is not to answer questions or to satisfy a need for information or to resolve an anomalous state of knowledge. The object of information retrieval systems is 'to help the users do these things'.[8] That is, the purpose of subject retrieval systems is to retrieve documents whose aboutnesses suggest that a user may find in them meaning(s) expedient to a certain need of the moment. Theoretical analysis of and justification for this assertion may be found in Cooper.[9-11] As Fairthorne[2] noted, however, we have been more engaged in deciding whether two documents are about the same subject than with determining the aboutness of a single document. The classificationist's conviction that it is both possible and useful to group documents according to their aboutnesses rests, nevertheless, on the assumption that the aboutness of a document is relatively fixed and stable, although its meaning(s) can and do change.

Most library and information science theories of document aboutness[6 8 12-14] are primarily theories of document meaning as the term is used here. In concentrating on obtaining access to documents according to the purpose for which a user wants them, these writers collectively assume that the aboutness of a document 'has been completely and correctly identified' by the indexer[15] and do not address the

problem of how an indexer is expected to make this correct identification. Other writers, however, have taken an interest in a document's aboutness as well as in its meaning(s). Hutchins,[16] for example, drawing upon the Czechoslovakian school of linguistics known as Functional Sentence Perspective (FSP), suggested that documents be indexed either according to their aboutness or according to their meaning(s), depending on the type of library and type of user for which the indexing is intended. Janoš,[17] expanding FSP techniques to full texts, concerned himself only with the aboutness that may be automatically extracted from scientific documents and assumed that a document with the correct aboutness will have the correct meaning for a user. Thus, the cognitive process of identifying the intrinsic aboutness of a document has not been systematically addressed, nor has a comprehensive theory of the aboutness of all documents been developed for use as a theoretical framework in library and information science research.

In her report on the CIG Colloquium, MacCafferty questioned whether the concept of aboutness had been adequately defined and asked, 'Is aboutness a function of a document, and if it is not, of what could it be a function? Is the aboutness of the whole document, its abstract, and its indexing all the same? Everyone seems to have his own idea as to what aboutness is.'[1] Likewise, classificationists creating subject classification systems have worked with varying concepts of a subject; Metcalfe[18] described the different treatments a number of classificationists have accorded the idea of 'subject'. In the present paper, the subject of a document is considered to be equivalent to its assumed relatively permanent aboutness, but not necessarily to its varying meaning(s), so that one may use the term 'aboutness analysis' for the cognitive process by which a classifier discerns the inherent at least partially independent topic, subject, or aboutness of a document. 'Aboutness analysis' is preferred to the traditional term 'subject analysis' because the latter has been used to describe both aboutness and meaning, but the term aboutness analysis preserves and emphasises this fundamental, although unproven, distinction.

The purposes of this paper are: 1, to explicate a potentially useful definition and theory of aboutness and aboutness analysis developed by text linguist T. A. van Dijk as part of his general theory of cognition and of text comprehension; 2, to explore some of the implications of text linguistics, particularly of van Dijk's work, for bibliographic classification theory; and 3, to suggest what elements a theory of the cognitive process of classifying documents would need in order to encompass and delineate how people identify the aboutness of documents.

In this way one may hope to respond, at least partially, to MacCafferty's requests for a framework within which aboutness can be considered, for an analysis of the extent to which aboutness is a function of a document and for a preliminary explication of the semantic relationship between the aboutness of a document, that of its abstract and that of its expression in a documentation language.

## 2. TEXT LINGUISTICS AND VAN DIJK'S THEORY OF ABOUTNESS

The general relevance of text linguistics for library and information science has been noted by Hutchins.[19,20] More recently Spector[21] suggested that we begin studying the analysis of multi-segment texts. Bell and Jones believe that indexers would benefit from using the techniques of 'macro-semantics: that is the study of large units of text'.[22] Van Dijk[23] outlined some of the intersections between his work and information retrieval research. Nevertheless, the predominant ques-

tions and findings of text linguistics have not been investigated extensively for their impact on bibliographic classification theory and practice. In general, text linguistics undertakes to describe and analyse a text as a complete coherent whole; this basic approach provides initial justification for exploring the implications of text linguistics for subject retrieval systems in general and for bibliographic classification theory and systems in particular because these research areas, too, characterise and thereafter treat documents as wholistic units. To explicate van Dijk's theory of aboutness and to examine some of its implications, however, it is first helpful to outline some of the general premises from which text theories arise.

Text linguists contend that the traditional theoretical unit of linguistic description and analysis, the sentence, cannot be used to explain the majority of linguistic behaviours. Instead, they posit that a different theoretical unit, the text, offers a more comprehensive and ultimately precise field of enquiry into language as people routinely use it. The bases for this conviction have been explained by, for example, de Beaugrande[24,25] and Brown and Yule.[26] A counter-argument that sentence grammars may adequately explain all linguistic behaviour appears in Morgan.[27] Derr[28] has outlined some of the implications for information retrieval systems research of recent cognitive science advances in the development of theoretical sentence grammars and semantics.

Text linguistics claims both spoken (e.g. lectures, conversations) and written (e.g. books, road signs, menus) texts or discourses of any length as its province. This paper is concerned only with written texts. A variety of distinctions can be made between a 'text' and a 'discourse', but for the purposes of this paper they may be treated as equivalents. A single word (e.g. 'Fire!') may be a text, but most texts are longer and correspond roughly to the intuitive notion of a completed topic-based unit of (written) linguistic behaviour.

Like transformational and generative sentence grammarians, text linguists accept the Chomskyan distinction between the deep and the surface structures of language. The deep elements of language are thought to be non-linguistic conceptualisations and cognitive integrations that are universal to human thinking but that are mapped onto the surface structure of a particular language in varied non-universal ways using an apparently unlimited array of linguistic devices. In addition, a single language may map the same deep logical concepts onto the available surface verbal elements in more than one way. For example, 'Mary threw the ball' and 'The ball was thrown by Mary' are generally considered to contain the same deep, but very different surface, structures.

In principle, the deep concept of a ball thrown by someone identified by a proper name is expressible with more or less precision in any human language, although the surface structure of the utterance can only be formed by means of the available surface options of the particular language and might, for example, necessitate a greater number of words, a different word order or different kinds of tenses from those available in English. In this view, the cognitive process of producing an utterance results from moving from a deep conceptual idea to a surface expression of that idea; and the cognitive process of understanding an utterance results from changing such a surface expression back into its deep underlying logical propositions. Any competent speaker of a language is assumed to be able to move smoothly and more or less unconsciously from deep to surface structures and back again and to be able, therefore, to create and reciprocally to understand well-formed sentences in the language.

Although text linguists accept these basic tenets of modern sentence grammars, they also contend that 'the text is *not* simply a larger "rank" than the sentence'.[29] That is, the sequence of sentences that make up a text are not mentally processed in a simple or linear fashion; the sentences in this paper cannot, for example, be treated as if they were logically connected only by 'and' or by punctuation marks. Such an assumption was made in some early transformational sentence grammars.[8 30] Further, text linguists contend that sentence grammars are overly concerned with the virtual language system, that is, with sentences that are theoretically possible, but that would probably never be created by a competent speaker of the language in the real world. A classic example is Chomsky's invention 'Colourless green ideas sleep furiously', which conforms to the rules of English grammar without normally conveying a meaningful message to competent speakers of English. In contrast, text linguists prefer to study the actual language system (i.e. all varieties of text) as people use it because actual texts reveal linguistic behaviour that has purposefully occurred as a result of the wish of a speaker or writer to communicate. Various kinds of intentionality and performative utterances have been studied by, for example, J. L. Austin.[31] Naturally-occurring texts (e.g. published documents) may thus be seen as the record of a human decision to act upon the choices offered by the virtual language system in order to exchange actual ideas, intentions, beliefs in real life.

The distinction between a virtual system and an actual system is similar to the distinction between a theoretical construct of any kind (e.g. the theory underlying a bibliographic classification system) and the concrete use of the construct in the world (e.g. the implementation of the classification system in a particular library), but it emphasises the reciprocal relationship between theory and practice without postulating a rigid boundary between the two. Like the distinction between aboutness and meaning, then, the distinction between virtual and actual language systems is based on intuitively attractive but unproven assumptions about an individual's relationships both to linguistic behaviour in general and to its manifestations in particular documents. In neither case is the distinction meant to imply any absolute isolation or conflict between the two extremes named by the terms chosen for the concepts.

To analyse the cognitive processes of both generating and comprehending actual language systems (i.e. texts or discourses) text linguists borrow from cognitive psychology the concepts of two kinds of mental information processing: 'top-down', concept-driven or deductive processing; and 'bottom-up', data-driven or inductive processing. Empirical research shows that these two kinds of mental processing appear to occur continuously and simultaneously during a reader's comprehension of a text.

## 2.1 Top-down processing
Understanding a written passage involves continual mental insertion into the text of extra-linguistic conventional knowledge of the world and of the general subject matter of the text. This conventional knowledge, brought from the reader's mind to bear on the text from the top down, is thought to be used to formulate and to test hypotheses regarding both the aboutness and the meaning of the text for that reader at the time of reading. This inserted knowledge may be common knowledge of the world (e.g. the steps involved in preparing to travel) and/or of the more specialised knowledge demanded by a particular text (e.g. theoretical physics). In all cases, however, such extra-textual knowledge is thought to be

cognitively organised for storage in memory in various kinds of large-scale systems known as frames, schemas, plans and scripts. These terms are used in differing ways, but for the purposes of this paper, extra-textual knowledge of all kinds is called 'frame knowledge'. The postulated differences among these cognitive structures are not needed here, and frame knowledge as a term implies an overall organising principle for all kinds of activities and ideas. Theories and models of frame knowledge have been developed for various kinds of artificial intelligence research and systems.

The indispensability of frame knowledge for comprehending texts has been shown experimentally. Non-determinate texts that offer only a few or a larger number of deliberately ambiguous clues about what frame knowledge is to be retrieved from memory and brought to bear on the text are treated by readers as nearly incomprehensible.[cg 32] One can intuitively grant, for example, that a knowledge of restaurants is necessary for understanding the invented text, 'The waitress came back and told me they were out of eggs. I gave up the idea of an omelette and ordered a ham sandwich instead.' Here, one must know that restaurants are places that cook and sell meals to customers, that a waitress is an intermediary between the customer and the kitchen, that omelettes are made from eggs and so on. A great deal of this kind of knowledge must routinely be supplied by the author of some kinds of texts (e.g. science fiction or anthropological field studies) because without an understanding of the customs and culture of the text world the reader becomes intellectually incapacitated. Usually, however, the author assumes that the reader possesses an appropriately detailed level of frame knowledge for the top-down processing of the text. The formal study of the knowledge an author assumes readers to possess already has been extensively pursued by FSP linguists.[cg 17]

One special kind of frame knowledge that readers bring to texts is their expectations for both the content and the structure of different kinds of texts. For example, we expect to find in academic papers precise bibliographic references to the work of other writers, but we would be surprised to encounter footnotes in a newspaper. In addition, certain kinds of overall structure are thought to be appropriate, for example, to narratives, but not to mathematical arguments, in which we do not expect chronological sequences, characters, dialogue or complications of plot. The study of readers' expectations for discourse content and structure is the study of text types or genres, but neither a typology of texts nor explicit text grammars for different text types is generally accepted. One writer, for example, identified eight text types: descriptive, narrative, argumentative, literary, poetic, scientific, didactic and conversational.[25] This is not the only typology that has been developed, and it is immediately apparent that these categories are not mutually exclusive, so this and other typologies are tentative at best. Much work has been done on story grammars[cg 33,34] and these story grammars have been found to differ in different cultures.[34] Van Dijk[35] has called the framework of expectations readers bring to different kinds of texts the 'superstructure' that is specific to a text type. Knowledge of the appropriate superstructure is, like frame knowledge, mentally inserted into the text from the top down during reading as an aid to immediate text comprehension and to eventual storage in and retrieval of the text from a reader's semantic memory.

*2.2 Bottom-up processing*
Van Dijk's theory of the bottom-up cognitive processing of texts provides a des-

criptive model of aboutness analysis. According to van Dijk, our understanding of a document's aboutness results from our ability to reduce the information in a text to manageable and therefore memorable proportions: 'our linguistic behaviour shows that we can say that a discourse, or part of it, was "about" something. That is, we are able to produce other discourses, or parts of discourses, expressing this "aboutness", EG in summaries, titles, conclusions or pronouncements in any form.'[36] Readers clearly do not remember a text word-for-word. They are, however, able to produce from memory a second (and secondary) discourse that expresses the aboutness of the first discourse because they can condense 'the full meaning of the text into its gist'.[37] The ability to restate the semantic aboutness of a discourse in this way originates in an automatic reductive cognitive process of summarisation that allows a reader to construct during reading a notion of the text topic and to store it in hierarchically-arranged memory structures for later recollection. Research evidence showing that representations of meaning are organised hierarchically in human memory has been summarised for the purposes of library and information science by Najarian.[38]

During the act of reading a text the reader notices the presentation of each sentence, automatically transforms its surface verbal structures into its deep conceptual propositions and establishes an understanding of the logical relationships between the words and sentences of the text. This process is called the microstructural analysis of the discourse, and it occurs on what has been variously called the local, sentence, or micro-level. At the same time, the reader engages in a global, textual or macro-level analysis of the text in order to arrive at an overall understanding of the aboutness and meaning of the complete text as a whole. In this analytic and synthesising process, sequences of sentences with their underlying logical propositions are cognitively compressed and summarised so as to produce a hierarchically-governed sequence of macropropositions (i.e. overriding logical propositions) that express the sum of the meanings of the propositions subsumed under them in the same way a series of co-ordinate sub-classes is assumed to be logically subsumed under its superordinate class. The end product of this continual text reduction is called the discourse topic or the aboutness of the text. Thus 'a concept or a conceptual structure (a proposition) may become a discourse topic if it HIERARCHICALLY ORGANIZES the conceptual (propositional) structure of the sequence.'[39]

These cognitive actions of compressing a text in order to generate a semantically accurate statement of discourse aboutness are, according to van Dijk, governed by macrorules that allow less important information to be dropped from memory during the reading and macroanalysis of the text. One formulation postulates five macrorules: the Weak Deletion Rule; the Strong Deletion Rule; the Zero Rule; the Generalisation Rule; and the Construction Rule. Brief examples of the postulated operation of each of these rules are given below. Throughout his writings van Dijk gives extended examples of the application of each rule both in words and in the notation of formal logic.

The two Deletion Rules direct the reader to select textual details that may be forgotten with impunity; in many cases attributive information does not contribute to the development of the main theme or macrostructure. For example, in the imaginary text, 'Mary played with a ball. The ball was blue', the reader may judge that the ball's colour is unimportant in the macrostructure of the text as a whole. Once the colour of the ball is forgotten, it cannot be retrieved inductively from memory because we cannot reason backward to reconstruct information that is not part of the conventional frame knowledge of playing with balls. The

Zero Rule is the opposing variant of the Deletion Rules; it admits a proposition directly into the macrostructure of the text with no reduction or summarisation. For example, in very short texts such as 'Come home', no information will be judged irrelevant.

The Generalisation Rule operates by mentally substituting the name of a superordinate class for instances of the class mentioned in the text in order to reduce the amount of detailed information the reader must remember. For example, 'Mary was playing with a fire engine, but she dropped it when Jane said, "Let's build a house". The two girls found the blocks and had started making a house when John came over and suggested going outside to play ball.' The Generalisation Rule is postulated to operate on this short imaginary text to summarise it for storage in memory as, for example, 'Children played with toys'. Here, two instances of the Generalisation Rule arise. 'Children' is the superordinate class for 'Mary', 'Jane', 'girls' and 'John'. 'Toys' is the superordinate class for 'fire engine', 'blocks' and 'ball'. Two things may be noted here. First, it is clear that frame knowledge for children at play is immediately activated to tell us that Jane means to build a toy, not a real, house and that the blocks are for building this toy house. Second, a different but equally adequate surface structure might also be used to express the generalisation that has been made (e.g. 'Some kids were playing with their playthings').

The postulated Construction Rule requires that a sequence of microlevel propositions be combined or integrated into one macrolevel proposition describing a complex event for which an extra-linguistic frame exists. For example, the imaginary text 'I brought wood, stones and concrete to my land. I dug a hole and laid a foundation. I erected walls, cut and fitted windows and doors and made a roof. Then I bought paint and painted my house' may be stored in memory as 'I built a house'. In this case, general frame knowledge allows us to recover through induction the various steps that would be likely to contribute to the macroproposition 'I built a house'. The information is recoverable because operation of the Construction Rule depends upon a frame that contains, in this case, the constituent materials and actions required for house-building. It should be noted that any details in the text that are not part of frame knowledge are exempt from the operation of the Construction Rule. For example, if the land turns out to be swampy and the house caves in, this textual component cannot be elicited from the frame knowledge for house-building.

In van Dijk's view, all five macrorules are applied continuously at appropriate places throughout the reading of the text and successive more or less unconscious applications of macrorules generate in the readers' memory shorter and shorter summarisations that express increasingly more general levels of macropropositions. In this view, the process of comprehension is a process of controlled forgetting. In a longer text than can be reasonably analysed here, there may be any number of summarised levels of abstraction that can be cognitively reduced to still higher levels. The final level, that is the level at which no further cognitive propositional reductions can fruitfully be made, is the topic of the text because it hierarchically organises all the detailed textual propositions into the most general macroproposition that meaningfully expresses the aboutness of the whole text. This recursive process eventually yields an expression of the aboutness of the text in the form of the text's highest appropriate macroproposition, which is then used to create a statement of the text topic in the surface verbal elements available in the particular natural language.

One of the automatic cognitive constraints that appears to operate upon this

recursion is that 'we will always stop at the lowest possible level of macro-structure, e.g. use the smallest superset or super-concept involved'.[40] This constraint means that, to return to the above example, a reader would not over-summarise the passage about the children at play as 'Somebody did something'. Instead, the smallest applicable generalisation will be used: 'Children played with toys'. This constraint guarantees that the final discourse topic will not be stated too broadly and that the aboutness of the text can be informatively remembered as its highest appropriate macroproposition while still retaining enough specificity to identify and to recall it when wanted.

Indeed, the highest macroproposition is 'essentially a title for that text unit'[41] in the same way that the name of a class in a bibliographic classification system is essentially a conceptual umbrella covering the ideas that are named in its subordinate classes. The class name thus resembles a generalised title and shares some of a title's expected characteristics. For example, 'One of the features of titles is the absence of verbal forms. Heavily weighted with nouns and adjectives, they offer the reassuring presence of a stable world, a static moment in the rush of time which the succession of words, pages, chapters, processed in linear fashion, emulates. Here, within the limited and isolated space of the title, we have access to the metalinguistic ground in terms of which we will be able to conceive the whole text. The reasonable assumption, we think, is that the title is a heuristic encapsulment, a point of departure as well as a point of arrival, a delimited and enclosed object.'[42] From this it seems to follow that the assumption that texts have a relatively permanent aboutness is mirrored in the static non-verb forms most often used to name the classes in a bibliographic classification system. Research in which subjects are asked to create title-like phrases for texts substantially confirms the view that readers expect titles to provide encapsulating concepts for the whole text.[43,44]

Van Dijk's theory of discourse processing thus provides a succinct and viable definition of textual aboutness and a potential model of aboutness analysis. One may say that the subject of a document is the highest specific macroproposition that is produced and can be expressed by a reader during cognitive reduction of a text by macroanalysis. Empirical research tends to confirm the general validity of this theoretical model.[eg. 33,34,45,46] The need for a theory of subject description has long been recognised in the literature of information retrieval systems, and, although the concepts of micro- and macro-levels of documents have been previously recognised,[eg 47,48] they have not been technically explicated in bibliographic classification theory. We may say, then, that van Dijk has formally described and analysed a cognitive process that can be assumed to operate during the aboutness analysis of a text for the purpose of classifying it by means of a particular classification system.

Although aboutness analysis is thus fundamentally a bottom-up inductive process of controlled forgetting, it should be kept in mind that top-down deductive insertion of remembered more or less detailed frame knowledge is absolutely indispensable for the comprehension of a text. To return to the above example, the reader must know that 'Mary' is the name of a girl, that girls are children, that children play with toys, that a toy house is intended and so on because without this extra-textual and contextual knowledge the postulated macrorules cannot operate effectively and the text is rendered meaningless. Many texts, of course, require much more specialised top-down frame knowledge for comprehension. Further, knowledge of the superstructure for a given text type is needed for an

understanding of a specific text; superstructures 'are, so to speak, the global ''form'' of the macrostructural ''content'''.[49] An intuitive appreciation of this has been apparent in bibliographic classification theory and practice at least to the extent of the inclusion of various kinds of form classes (e.g. encyclopedias, biography) in most classification systems. In addition, as previously noted, some work has been done on delineating particular text types for the purpose of aboutness analysis.[e.g. 17]

Van Dijk's concepts of superstructures and macrostructures offer a relatively specific formulation of how people, presumably including classifiers, mentally formulate the aboutness of written texts. These concepts might be fruitfully adapted to an investigation of the cognitive processes of classifying documents. Tentative descriptions and explications of 1, the relationships between the classified documents themselves; 2, the relationships between classification systems and classified documents; and 3, the cognitive process of classifying documents can be drawn from the concepts of top-down and bottom-up cognitive processing and from van Dijk's theory of aboutness identification for text comprehension and storage in semantic memory.

### 3. INTERTEXTUALITY AND CLASSIFICATION SYSTEMS

The aboutness of a written document, then, may be postulated to be a function of both its internal content micro- and macrostructures and its external relationship(s), one of which is its superstructure. This generalisation can be diagrammed[50] as in Figure 1.
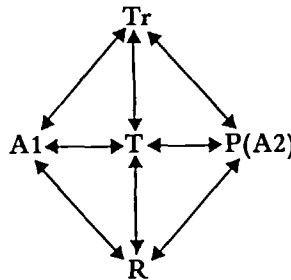


*Explanations*

A1   : Author (of the prototext)
T     : Text
P(A2) : Percipient (author of the metatext)
R     : Reality
Tr    : Tradition

FIGURE 1. *Relationships of a text*

Here, the text (T) mediates between and interacts with four extra-textual elements: cultural tradition (Tr), including readers' learned general expectations for the superstructures of various text types; the reality (R) of the moment, including the reason the reader is interested in and drawn to the text; the author of the original (proto)text (A1); and the percipient or reader (P). The percipient may in turn be an author (A2) of a (meta)text, 'such as criticism, digest, epilogue, essay, text interpretation'.[50] In van Dijk's terms, the metatext obtained by a

reader's macroanalysis not only expresses the aboutness of the prototext, but is a fully new text in its own right. He writes, 'A summary, then, will be taken as a discourse expressing a macrostructure of another discourse.'[51] Nevertheless, there is clearly a relationship between a text and its derivative metatexts. This relationship has been called intertextuality.

Intertextuality is 'the principle whereby the textuality of any one text arises from interaction with other texts'[52] and 'subsumes the relationships between a given text and other relevant texts encountered in prior experience, with or without mediation'.[53] The existence of various forms of intertextuality is unquestionable and has been consistently recognised in the traditional cataloguing and classification practices, to cite one example, of indicating the edition of a work in its cataloguing and of cuttering these different editions so that they come together on the shelves. Fairthorne's identification of the 'classificatory landscape'[54] raised questions of intertextuality for discussion among bibliographic classification theorists. Some of the general issues that recognition of the concept of intertextuality raises for readers, writers and critics are treated provocatively from a philosophical and linguistic perspective by Steiner.[55,56] The differing points of view of various writers are expressed, too, in Valdés and Miller.[7]

One set of intertextual relationships exists between a work, its various derivative metatexts such as its summary or abstract and its expression in a documentation language. We may call the original text the 'primary text', a summary or abstract the 'secondary text', and the expression of the primary text in a documentation language the 'tertiary text'. These three texts, although all ideally express the same conceptual aboutness, are considered to be in addition independent texts in their own rights. The surface verbal expressions in a derivative text may and probably will be different from those in the original primary text. Fairthorne,[57] for example, pointed out the textual status of derivative texts by calling thesaurus terms 'minimal texts'. Thought has also been given to ensuring that an abstract reproduces the aboutness of its primary text,[e.g. 58] and work on standardising vocabulary so that the same deep concepts may be expressed in the same words, is carried out, for example, under the auspices of Infoterm.[e.g. 59]

An approach to intertextuality that focuses more narrowly on bibliographic classification theory and practice may distinguish two basic but not mutually exclusive kinds of intertextuality. First is the intertextuality that obtains between the primary texts of documents that, by virtue of having been assigned to the same class in the same classification system, are intertextually related to one another. Second is the intertextuality that obtains between a bibliographic classification system as a naturally-occurring primary written text in its own right and the primary documents that have been classified under its authority. In this sense to classify a document is to assign it to the class whose name can be seen as a suitable derivative tertiary text (e.g. a potential title) for the document.

Text linguists and other text researchers have advanced various criteria for defining and determining the existence of textuality. For the purposes of this paper we may grant that the written text of a bibliographic classification system falls within the definitions proposed for texts.[e.g. 60] A classification system may therefore be analysed with the same concepts that are used to analyse other primary texts. In this paper the entire text of a bibliographic classification system – including its introduction, instructions, schedules, users' manual and anything else meant to be an official adjunct to it – will be called 'the classification' or 'the classification text'. In order, however, to avoid an additionally complicating level

94

of intertextuality, secondary works of evaluation, interpretation, review and/or history of the classification system are excluded. In this way, bibliographic classification systems in general may be treated as a text type, so the discussion now primarily concerns a description of the paradigmatic form that modern bibliographic classification systems have commonly assumed. Of course, any actual classification system may conform in its details and overall character only more or less strictly to the characteristics of the classificatory text type, just as the report of a scientific investigation may not mirror perfectly all the superstructural expectations readers have for its text type. Nevertheless, it is helpful first to characterise classification texts as a general text type by isolating and describing two basic kinds of intertextuality. Such an analysis of the relationships 1, between classified documents and 2, between classification systems and the documents classified by them may identify elements and complexities that need to be accounted for in any model that hopes to sharpen our focus on the cognitive act of classifying.

*3.1 Intertextuality between documents classified in the same class of the same classification system*

In general, the aim of bibliographic classification systems has been to group documents on the basis of certain presumed helpful subject likenesses and, as a by-product, to separate those same documents on the basis of certain presumed less helpful subject differences. To the extent that this goal is a realistic one, primary documents that can be appropriately classified in the same class of the same classification system are intertextually related to each other on the basis of their shared subject(s). That is, the highest macroproposition of each document is subsumable under the same class name in the classification system; therefore, primary documents that can be reasonably placed in the class share similar aboutnesses and are intertextually related to each other in varying degrees for this reason. Depending upon the conceptual breadth described for similar classes in different systems, however, a single document may be found to partake of different intertextual relationships when it is classified by means of different classification systems.

The assertion that subsumability under the same class name creates an intertextual relationship between documents relies upon one of the fundamental notions that Euclid identified as common to all of science; that is, 'things that are equal to the same thing are also equal to each other'. Here, we may say, admittedly with less precision, that 'things (i.e. documents) that are related to the same thing (i.e. a certain class name in a certain classification system) are also related to each other'. The same kind of assumption is made in the literature of citation analysis where the assertion underlying studies of bibliographic coupling is that 'things that cite the same things are also related to each other' and the assertion underlying studies of cocitation is that 'things that are cited by the same thing are also related to each other'.

The assumption that documents classified in the same class share subject-based relationships is possibly the most fundamental rationale for creating a system that will group documents by subject. Blair suggested that 'subject descriptions or the identification of topics are heuristics which supplement a more basic skill which the enquirer makes use of – his ability to make judgements of similarity.'[61] Blair treated the skills of the user of the system, not the skills of those who are classifying documents for the user's eventual benefit, but he correctly noted that the mental ability to judge similarities is 'at present, unanalysed',[62] although some

kinds of similarity judgements have been studied in cognitive psychology.[68][63] It can be questioned whether the ability to judge aboutness similarity between documents is a 'more basic skill' than identifying the aboutness of one document because it must be difficult to judge the sameness of things without first noticing the characterising elements of one thing alone. Nevertheless, the conviction that judgements both of document aboutness and of aboutness similarity can rationally be made lies at the heart of the attempt to create bibliographic classification systems and to use them as analytic tools for classifying primary texts by subject.

In a discussion of the nature of subject fields, Dahlberg peripherally addressed the question of intertextual relationships between documents classified in the same class. Basing her analysis on the work of Diemer, she stated that 'a science . . . may be regarded as a system of propositions on a certain area' so that 'the defining name of a science may then be regarded as the hierarchically highest-level proposition from among the entire system of propositions forming that one science.'[64] Dahlberg's point was carried a step further by van Dijk, who suggested that hierarchically-related propositions both state the aboutness of documents and contribute the name for the class that can contain those documents in an information retrieval system. Van Dijk noted 'sets of summaries [i.e. groups of secondary aboutness-expressions] may in turn be subject to further organisation, on the basis of their underlying propositions (which are macro-propositions of the corresponding discourses).'[65]

Such further organisation, according to van Dijk, generates a system in which overriding concepts are expressed in a descriptor language. 'The concept expressed by a descriptor is, thus, a function characterising a set of summaries, namely, the set of summaries which use the concept in their macro-structure'; and yet, 'descriptor sequences cannot possibly account for the (macro-)meaning of single discourses; they only define sets of summaries'.[66] That is, the name of a class in a classification system delimits the allowable aboutness range for primary documents that may reasonably be assigned to the class, but the name of the class does not undertake to make a detailed statement of the aboutness (or of the meaning(s)) of individual documents. Bliss expressed essentially the same idea when he wrote that a 'class is *potentially* complete, comprising not only existent but *all* past and future or possible things that may be defined by its definition or named by its name.'[67]

It follows from these views that the necessary similarity of the highest appropriate macropropositions of all the documents that can be grouped under a certain class name connects each document intertextually to all the others. Conversely, the connected macropropositions of the documents in the class can be cognitively reduced to the name of the class so that, as van Dijk and Bliss pointed out in their different ways, the name of a class in a classification system defines a set of summaries at least to the extent of providing a generalised title for all the documents classified in any particular class of the system. A document whose summary is not defined by the class name does not, therefore, belong in the class. Thus, an automatically-occurring intertextual aboutness relationship between documents appropriately classified in the same class of the same classification system can be identified and needs to be taken into account.

It should be noted, however, that this theoretical statement does not address the difficult practical problem of deciding whether the surface expression of a concept in a class name is an adequate verbal representation of the underlying deep proposition(s) that may be extracted from the document. This decision is one that

classifiers of documents must make during the classifying process. To return to the terms previously suggested, we may say that the document itself expresses its own primary aboutness and that the name of the class in the classification system is a tertiary expression of the aboutness of the primary document. The mediating secondary aboutness expression for the document is generated by the classifier while analysing the document for the purposes of classifying it. This secondary aboutness description of the document may be made explicit in writing, as in an abstract or summary. In most cases of classifying, however, a classifier relies on the automatic cognitive processes described in van Dijk's theory of aboutness analysis and does not write an explicit secondary text before classifying the document.

*3.2 Intertextuality between a classification system and the documents classified by it*
Intertextual linkages between a classification system that is expressed in writing as a bibliographic classification text and the documents it is used to classify have not been fully examined in the literature of classification theory. A preliminary exploration of this relationship may offer insights into the cognitive processes of the classifier, but it is first fruitful to explore more explicitly some general characteristics of bibliographic classification systems as a text type and to describe classification texts with the analytic concepts offered by text linguistics in general and by van Dijk's theory of aboutness in particular.

Like all texts, a classification system has a highest appropriate macroproposition that informs the system as a whole and that states the aboutness and the conceptual argument of the entire classification system. General classification systems are said to be about the whole world of knowledge, while special systems undertake to organise a smaller knowledge base. In either case, the highest appropriate macroproposition, i.e. the aboutness of the system, is rooted in the cultural and bibliographic heritage that produced it and that also provides the extra-linguistic and extra-classificatory frame knowledge that a classifier must bring to the process of classifying documents. This assertion was diagrammed in Figure 1. Lee[68] suggested that these elements of classification systems be called the 'cultural warrant' of the system. Recognition of the existence and significance of such a cultural warrant for bibliographic classification systems is implicit in works[8,69] that formulate methods of extracting a sociology of knowledge from the uses made of documents classified by a particular system.

Interest exists in whether a classification system can be universally applicable and valid. Classificationists have usually hoped that their particular creation would be more or less permanent and to this end have tried to find unaltering bases for their systems. Bliss, for example, founded his Bibliographic Classification (BC)[70] in contemporaneous philosophical notions of the system of the sciences in the belief that this system 'remains stable and permanent.'[71] Ranganathan believed that the methodology of the Colon Classification (CC) would only 'occasionally' be superseded by 'new developments in the world of knowledge'.[72] Dahlberg[73] proposed a classification based upon root concepts rather than on the academic disciplines. D. Austin[74] believed that there may be a universal syntax, but not a universal content, for information retrieval systems. The various possibilities for and definitions of universal classification have been collectively explored at an international conference.[75]

As D. Austin[74] for example, recognised, culturally-determined components of classification systems seem difficult, if not impossible, to dispel. The capitalistic

bias of the Dewey Decimal Classification (DDC) was recognised in the USSR, and, after the death of Stalin's wife, Nadezhda Krupskaya, in the 1940s, a new classification was developed that would not be 'bourgeois' in 'conception'. This anecdote 'illustrates one practical case in which the implicit hypnotic power of book-classificational arrays was intuitively recognised'.[76] Others[8][77] have noted various biases in DDC, and attempts to modify or remove them have been ap- plauded.[8][78] The difficulties encountered in translating a classification system into another natural language, too, testify to the culturally-determined underpinnings of bibliographic classification systems.

In a series of lectures delivered at Cambridge University, Hulme[79] advised historians to study bibliographic classification systems as cultural indicators because, he argued, statistical analyses of the literary warrant expressed through a classification system would illuminate the thought and structure of the society for which that particular system was especially devised. This argument can be seen as a development of Hulme's concept of literary warrant.[80] Hulme[79] proceeded to make preliminary analyses of the literatures of architecture and of textile in- dustries during the Middle Ages and of patent applications in England to show how his thesis might be extended. This application, which Hulme called 'statistical bibliography', is considered to mark the genesis of modern studies of citation analysis and bibliometrics.[81] Recently de Grolier carried Hulme's line of enquiry further by finding in bibliographic classification systems evidence of cultural evolution. Calling his thesis 'rather sketchy',[82] de Grolier nevertheless argues that careful elaboration of his methods would probably reveal not only the cultural orientation of the relevant society but also the individual interests and opinions of the particular classificationist.

Whether or not a universally valid syntactics (e.g. citation order) and/or semantics (e.g. non-culturally orientated warrants for classes) is possible for a bibliographic classification system, neither has been universally accepted in theory nor adopted in practice. We may then posit the cultural and bibliographic heritage of the classification system – its cultural warrant – as the highest macropropositional expression of its aboutness. In refraining from stating the aboutness of the classification system as 'knowledge', we respect van Dijk's previously-discussed constraint that operates during aboutness analysis and postulate that the aboutness of a classification system is the smallest macroproposi- tional superset that will both contain and identify it. This aboutness might be ex- pressed, for example, as 'the organisation of the knowledge of this subject(s) at this time and this place'. This cultural and bibliographic heritage would include all the assumptions, beliefs and traditions that exist in the culture in question and also the specific elements of theoretical and practical classificatory tradition, of the way in which the particular classification system works and of familiarity with the collection to which the document is being added. A classifier examining a docu- ment will of necessity bring these cultural and classificatory frames to bear both on the classification system and on the document from the top down in the same way a reader brings frame knowlege to the comprehension of any text.

In possessing a highest appropriate macroposition and in requiring the insertion of top-down frame knowledge by the user, a classification text is like any other text. In contrast to other texts, however, a classification text has a unique addi- tional purpose; to provide named slots to accommodate conceptually and to ex- press notationally the topics of other texts. Other naturally-occurring but non- classificatory texts focus on leading the reader to a comprehension of their own

aboutness, but a classification text focuses on serving as an organising principle and expressive tool for encapsulating the aboutnesses of other primary documents, including other classification texts.

To attain the end of accommodating the subjects of other primary texts, a classificationist creates a primary text of named conceptual slots so that the classification system can provide a superstructural form into which, following the instructions for the particular system, the classifier may insert the macrostructural content of the other primary documents undergoing classification. A classificationist, then, consciously endeavours to design an artificial virtual system that can be used to organise and to express the actual systems of other naturally-occurring texts. In this, the classificationist develops a virtual system, that, like a natural language, can be manipulated by the classifier to express deep conceptual propositions in the available surface structures of the natural language in use (e.g. current terminologies) and of the classification system itself (i.e. its notation).

At the same time, however, the classification text is itself a naturally-occurring actual language system written in a natural language. As such it is therefore subject, as we have seen, to cultural orientations and to the vagaries of expression in the available surface structures of a natural language. The problems of consistently expressing deep concepts in the surface structures of natural language are familiar ones in the literature of information retrieval systems. These problems are responsible, for example, for Ranganathan's complaint about the 'immaturity' of natural language,[83] for efforts to standardise terminologies, for the existence of controlled descriptor vocabularies of all kinds and for the need to develop switching languages between these vocabularies. There is no reason to suppose that similar problems and potential anomalies do not exist in naming the classes in a classification system.

Thus, the ideal classification text creates a virtual system by means of an actual system and is in this sense the opposite of a natural language, in which a virtual language system is the means by which actual language systems (i.e. naturally-occurring discourses) are created. In natural languages the underlying virtual language system is used to create texts, but in a bibliographic classificatory language the artificial virtual system cannot serve to generate completely new actual systems (i.e. naturally-occurring texts). Instead, it can only provide named conceptual slots and notational devices hospitable to the subjects of existing or potential texts. As has been noted, the names of classes in a bibliographic classification system can be seen as derivative texts in their own right; nevertheless, one important feature of these tertiary texts is that they are always shortened, because cognitively reduced, versions of the aboutness of a primary text. For this reason, a classification system cannot be used to create original texts, but must be reflective and reductive in nature. From this perspective, attempts to increase notational hospitality and expressiveness by means of facet analysis and synthetic notations may be seen as attempts to increase our ability to use the classification system to create new texts. Although information retrieval systems are routinely called 'languages', there are significant differences between natural and logical languages.[8, 21] One of the fundamental differences between natural and classificatory languages is pinpointed by the distinction between virtual and actual systems and by highlighting the superior text-generating capabilities of a naturally-occurring virtual system, e.g. a natural language.

We have introduced, however, a more refined version of this distinction than we have used before. This new distinction exists between an artificial virtual

system (e.g. a bibliographic classification system text) and a naturally-occurring virtual system (e.g. the deep structures of natural language). Research is needed to determine more precisely which characteristics these two kinds of virtual system do and do not share. It seems to follow, however, that the act of expressing the aboutness of a primary document in the surface structures of either kind of virtual system would call upon similar cognitive structures and would call for similar cognitive reductive reactions to information. It also seems to follow that, just as all sentences in English may be said to be related to each other by virtue of the very fact that they partake of the surface structures of English, so too primary documents classified by the codified artificial virtual system of a classification text are related to that text because it has been used to limit and to structure the allowable (notational) tertiary expressions of their aboutness. That is, notational expressions of document aboutness in a classification system are related to that system in the same way and for the same reasons that sentences *in* the English language partake *of* the English language.

The general characteristics of the classificatory paradigm or text type, then, are:

1. *A classification system, as expressed in a classification text, possesses an in-* dependent aboutness of its own that, as in the case of other naturally-occurring texts, can be comprehended only with the help of extra-textual and extra-classificatory frame knowledge; and

2. A classification system, as expressed in a classification text, attempts through its introduction, instructions, schedules, index and/or other official adjuncts to codify an artificial virtual system for the expression of the macropropositional aboutness of other actual texts that have arisen through a naturally-occurring virtual system (i.e. a natural language).

With this analysis of the two kinds of intertextuality inherent in classification systems and of the general classificatory text type in mind, it is possible to discuss more precisely the theoretical cognitive processes of classifying documents. The classifying process can be seen as a secondary process that mediates between recognition of the primary aboutness of the document-to-be-classified and the tertiary expression of its aboutness in a named class by means of its corresponding notation in a particular bibliographic classification system.

## 4. THE COGNITIVE PROCESS OF CLASSIFYING DOCUMENTS

### 4.1 A theoretical model of classifying documents

Library educators agree that students must learn to classify documents, and the term 'translating' is sometimes used for the process of changing the classifier's perception of the aboutness of a document from a natural language surface expression into an appropriate classificatory notational surface expression. The questions then arise: What is one learning to do when one learns to classify documents? What cognitive processes operate in classifying? The idea of translation from one natural language into another does not seem complex enough to account for the process of classifying because, as we have discussed, a classification system, although written in an actual natural language, represents in addition a different consciously and artfully developed virtual language system whose use must be learned and continuously and consciously applied to the document by the classifier. The differences between using a natural and a classificatory language and the additional problems these differences create in the translation process,

then, must be accounted for in a theoretical description of the cognitive process of classifying documents by means of a particular classification system.

In van Dijk's theory of textual aboutness the reader processes a text from the bottom up and simultaneously inserts pre-existing top down frame knowledge into the reading of the text. These complementary processes are thought to result in an understanding of what the text is about. Other elements may, however, enter into a reader's comprehension of a text. Comprehension may, for example, 'be influenced by other cognitive factors, such as interest, task, purpose, knowledge, norms or opinions and attitudes'.[84] In reading for a special reason, 'the special purpose overrides whatever text structure there is'[85] and is thought to skew the reader's understanding of the text. Empirical research tends to uphold this view. For example, Black found that 'the inferences people make while reading a text are affected by the purpose for which the text is read'.[86]

This element in the comprehension of texts is usually called the 'interpretation' of the text[87,88] and consists, in general, of reading the text not in a 'neutral' fashion with the purpose of simple comprehension, but with a directed purpose such as deciding whether one agrees with it or whether it is interesting, among many others. Discussions of interpretation usually centre around the insertion of various kinds of frame knowledge directly from the reader's mind;[8,86] but we are concerned here with the process of cognitively reconciling two *written* texts, the classification text and another primary document, in order that the aboutness of the one can ultimately be expressed in the notational system of the other. Evidently, cognitive integration of the intertextual relationships of two written texts (one in an artificial and one in a natural virtual language system) in exactly this way has not been studied. Ranganathan's concepts of the 'idea', 'verbal' and 'notational' planes of classification systems seem to be theoretical terms recognising the potential difficulties of integrating two texts from different modes of expression.

Figure 2, modified from Figure 1, after Zsilka[50] shows a schematic representation of what a theory of the cognitive process of classifying documents needs to describe and account for. Since van Dijk postulates that a special purpose overrides text structure during comprehension, Figure 2 shows the classification text superimposed upon the document text being classified. In this, the aboutness of the actual classification text (i.e. the world of knowledge it seeks to organise) and the frame knowledge (e.g. classificatory tradition) needed for its comprehension and use take precedence over the aboutness of the primary document and the frame knowledge needed for its comprehension. The percipient of both documents is the classifier, who must merge the aboutness and extra-textual frame knowledge of the classification system with those of the document by following the instructions for the use of that system. Further, in order to preserve and to express continuing intertextual relationships between documents classified by means of the particular system, the classifier must endeavour to merge the two texts in the same way each time documents with the same or sufficiently similar highest appropriate macropropositions are classified. As Figure 2 shows, extra-linguistic reality (e.g. the particular library for which the document is being classified and literary warrant at the time of classification) adds a further ingredient that must be accounted for in a theoretical description of the classifying process.

We may note then that classificatory expression of the document's aboutness (which we have called a tertiary derivative text) dominates the original or primary document text. The primary document, which is the one that interests users,
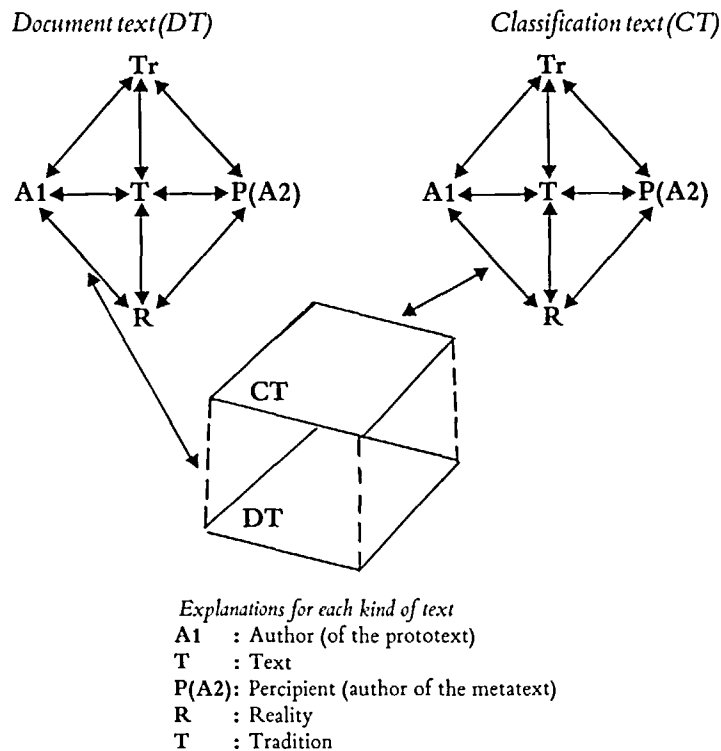
101

FIGURE 2. *Intertextual relationships of classification text and document text*

must yield to the exigencies of the classification system. For this reason it is impossible for the classification system to respond quickly to changes of literary warrant; recognition of the time lag between primary and derivative texts results in complaints about the difficulty of using an outmoded classification text to classify contemporary documents. In this sense, the primary document may be seen as a passive recipient of active analytic treatment by a classification system that may ironically (for to do this well is its stated purpose) be unable to deal adequately with the document's macropropositional aboutness.

Thus, a theory of the cognitive process of classifying documents must explain how the classifier sequentially:

1, transforms the surface structure of the document into its deep propositional logical structure;

2, transforms the surface structure of the classification system into its deep propositional logical structure;

3, joins these two deep propositional structures using the actual system of the classification schedules and instructions as an artificially-constructed virtual system for expressing the natural language actual systems of primary documents; and

4, transforms the resulting single propositional structure of 'classification system applied to document' back into the surface structure of the classification schedules, generates the apposite notation, and concludes that the document has been appropriately placed in the class with other documents to which it is most

nearly intellectually and intertextually similar. This general statement of the cognitive process of classifying documents covers what one learns to do when one learns to classify.

The first two steps in this process are described in Chomsky's sentence and in van Dijk's text theories. There seems no particular reason to believe that, given knowledge of the appropriate superstructures and frames, the act of comprehending one text (e.g. a classification text) is fundamentally different from the act of comprehending another (e.g. a document text). Thus, the previously-cited results of empirical research into reading comprehension can be assumed to be at least generally valid for both types of text. The problem seems to arise in the third and fourth steps above, when the classifier must merge the deep propositions of two written texts according to the instructions and conventions of one of them and then generate a single surface notational expression of the fused deep structures. It is these two steps that are often seen as a translation of a natural language expression into a classificatory language expression. By successfully completing these steps the classifier establishes intertextual relationships between the classification system and the documents classified by it and between documents that appropriately belong in the same class of the same classification system.

Lancaster, for example, divided the classifying process into two intellectual steps: 'the conceptual analysis . . . of a document, and the translation of this conceptual analysis into a particular vocabulary. It is rare that these two steps are clearly distinguished.'[89] Lancaster's recognition of distinct stages in classifying is valuable. His formulation, however, does not strongly contrast step two above from steps three and four and so glosses over the mental procedure of consistently integrating the frame knowledge, superstructures and macrostructures of a classification text with those of a document text before a surface expression in the classificatory notational language can be generated. Evidently these processes have not been completely isolated, but with the addition of further concepts from van Dijk's work, it may be possible to sketch the overall design of a research project that could shed light on some of the issues raised by the identification of intertextual links between a classification system and the documents classified by it and by the resulting descriptive model of the cognitive processes of document classification as shown in Figure 2.

### 4.2 Studying the classifying process
Van Dijk[6] distinguishes two kinds of mental relevance assignment readers make during reading comprehension. First is normal textual relevance assignment, which occurs at either the micro- (sentence-) or the macro- (text-) level of analysis. At both levels normal relevance assignment is distinguished from contrastive or differential relevance assignment. At the macro-level, normal relevance is the independent aboutness of a discourse and differential relevance is the basis for what we have called the meaning of the text for a user. Differential relevance assignment occurs when one 'differentially selects items for ''special teatment'' from among *similar* items (i.e. items on the same [hierarchical] level)'.[90] Such special treatment is thought to result from greater interest, an immediate purpose or a special task, among others. If, for example, it is for some reason important to the reader to remember that the children were playing with blocks, 'blocks' would not be treated to the same completely reductive generalisation process as were the other toys. The topic of the passage might then be stored and remembered as 'Children played with toys, some of which were blocks'.

Second is contextual relevance, in which the 'cognitive (and social, communicative) context defines what elements of the text are found important by the reader'.[91] In the case of classifying documents, these elements of contextual relevance would be contributed by the cultural and classificatory warrant of the classification system and by the particular library for which the classifying is done.

This paper has primarily treated aboutness and meaning in documents. Aboutness in a document is what the classifier determines during aboutness analysis and meaning in the document as the reason a user may want to retrieve it. In addition, however, the classification text, like other texts, may be assumed to have on its own behalf both an aboutness and a meaning for the classifier; who must ascertain the aboutness of the document by means of the aboutness and the meaning of the classification system as in Figure 2. The classifier, then, knowing that the document must be classified with a particular classification text, may attach undue weight to conceptual elements of the document that can be dealt with easily in the context of the particular system and devalue conceptual elements that are less easily accommodated (or not accommodated at all) by the system in use.

In this case the conceptual relevance of the aboutness and meaning of the classification system influences the conceptual analysis of the document and might produce an aboutness assessment differentially related to the available structure, classes and notational devices of the classification. If differential relevance assessments were to vary, due, for example, to the differing conceptual complexity of different documents or to different levels of frame knowledge possessed by different classifiers, inconsistent classifying could result. Although Rolling[92] shows that measures of consistency do not reveal the quality or effectiveness of indexing, consistent aboutness analysis has been deemed crucial for the efficient retrieval of intertextually-related documents from information systems.[68.93] It is thus justifiable to seek possible causes of inconsistent aboutness analysis in the hope that knowledge of causes might lead to methods of lessening inconsistencies or anomalies in document analysis.

In discussing the classificatory superstructure or text type, frame knowledge necessary for understanding the classification text has been treated only paradigmatically. Although all classification systems have as their aboutness a certain general or special world of knowledge at a certain time and place, one system has different specific frames from another: different overall structure, classes, synthesising possibilities, standard citation orders, to name a few. Moeller compared classifications of 180 monographs by DDC as used in the British National Bibliography (BNB) and by the State Library Classification (SLC) used in Denmark. He found that 'in most cases the two [classification] codes, being universal frames of reference, interpreted the topic of the same book differently'[94] and that responsibility for this difference lay in the very different structures of the two systems. Thus, although there is a paradigmatic classificatory superstructure, text type or genre, the use of a specific classification system as an analytic tool produces a different aboutness analysis for a document in the same way that one novel, although partaking of an overall narrative superstructure, differs markedly from other novels and presents the reader with a unique view of the world. As previously noted, conceptual breadths of similar classes in different systems may result in the isolation of different intertextual relationships between the documents classified.

It follows that the aboutness and meaning of a particular classification system and its contextual relevance may influence a classifier's aboutness analysis of a

primary document. A classifier, knowing that the document's highest appropriate macroproposition must be slotted into the available macropropositions (i.e. the named classes) of the classification system, may unknowingly fail to analyse the document independently from his/her knowledge of that particular classification system. This possibility means that documents with similar aboutnesses may be classified differently by different classifiers or by the same classifier at different times. Some of the findings of the Cranfield retrieval tests seem to suggest this result.[95] It would be useful to know to what extent and in what specific form(s) differential relevance assignment presents itself during the aboutness analysis of primary documents. Empirical research may one day begin to answer this question.

Such research might begin by testing the model of classifying documents presented in Figure 2. An experiment could be devised that, on the premise that readers assign normal relevance to textual elements when reading for simple comprehension, tested whether expecting to have to classify a document by a particular classification system significantly changed a classifier's perception of its aboutness and caused the assignment of differential relevance to certain textual elements. Previous studies[e.g. 34,37,45] have shown that people produce remarkably (conceptually-)uniform protocols when asked to summarise materials in writing. In principle, subjects who do not expect to classify a document should produce equally uniform summaries. According to the analysis of the cognitive process of classifying in Figure 2, however, the need to merge the deep and surface structures, superstructures and frame knowledge of two written texts may override the independent macroanalysis of a document if a classifier expects to have to classify it later using a certain classification system.

This prediction might be tested with two groups (e.g. students or experienced classifiers) who have roughly the same amount of experience in classifying with a certain classification system. Both groups would be given the same texts. One group would be first instructed to read and summarise the texts from memory with no reference made to classifying them. After the summaries were written, this group would be asked to classify the texts using the familiar classification scheme. The second group of participants would be given the same instructions to read and to summarise from memory, but would also be told beforehand that they would classify the texts later with the familiar system. The classification text to be used would be available to this group while the summaries were being written, and subjects would be told to think about how to classify the texts while reading and summarising. Afterwards, the two sets of summaries and notations could be compared.

Under the terms of van Dijk's theory, we might hypothesise that the first group, who had summarised the texts without expecting to classify them, would reproduce the macrostructures of the texts with normal relevance assignment. Conversely, the group that anticipated classifying the texts could be expected to assign differential relevance to certain parts of the text. This differential relevance assignment should show up in deep conceptual differences in the summaries and notations produced by the two groups. That is, the group that expected to classify the document should show a statistically significant higher rate of differential relevance assignment for certain elements in the text, depending on which textual elements could be easily classified in the particular system. In addition, the two sets of notations could be compared to see whether a greater variety of notations had come from either group or whether, even though the first group had summa-

105

rised the texts without expecting to classify them, they arrived eventually at the same notations as the second group. Comparisons of the two sets of summaries and notations might thus suggest how and to what extent aboutness analysis is or is not influenced by the classification text.

At least two major problems arise in the design of a study of this kind. First, in addition to the necessity of considering the traditional kinds of internal and external experimental validity and invalidity,[e g 96] it is often thought that test situations should reproduce as closely as possible the conditions under which participants might normally do whatever the test entails.[cf.97] This attempt to produce nearly-normal test conditions is sometimes called establishing 'ecological validity' for an experiment. An ecologically valid classifying situation would ideally include several elements. For example, classifiers normally have access to both the classification schedules and the primary document and can refer to either as necessary during the classifying process. Thus the group that did not expect to classify the document and was not given access to the classification text while reading the document would be at an ecological disadvantage during the experiment. In addition, both groups would be summarising the document from memory, so that normal access to the document would be denied them during the summarising stage.

It can also be argued that making participants in both groups read the whole text would decrease the ecological validity of the experiment. Classifiers do not usually read an entire document. It is generally asserted in the literature of classification theory and practice[e g 98] that an examination of the table of contents, preface, chapter sub-divisions, indexes and other parts of the document usually suffices for classifying it. Reder and Anderson[99,100] have shown, however, that memory for summaries is greater than memory for the texts themselves, and Tell[101] found that indexing consistency dropped when indexers used full texts rather than abstracts or titles. Thus, reading a whole text and summarising it before classifying might have a strong influence on the classificatory decisions and efficiency of both groups.

The second major problem that would have to be overcome before an experiment of this kind could be conducted is one that has been faced by all experimenters in text comprehension studies. There is no generally accepted manual or automatic procedure for determining whether the deep structure of the discourse has been adequately expressed in the surface structure of a natural language. People can legitimately express the same idea in a number of equally adequate ways. Thus, in experiments of this kind a prototype summary that is judged to contain normal relevance assignment for each part of the text is usually prepared. This summary is often written by the experimenter and submitted to judges for verification of its conceptual accuracy.[e g 99,100,102] After the experiment, a way must be found to decide whether and to what extent the summaries produced by participants conform conceptually to the prototype summary. Usually, experimenters personally compare participants' individual summaries to the conceptual deep structure (but not to the verbal surface structures) of the prototype summary.[e g 103] In this case, of course, an experimenter's unconscious bias may seriously compromise experimental results.

No entirely satisfactory solution to these problems apparently exists. An experiment to test Figure 2 might require considerably more detailed analysis of the problems of ecological validity and of experimental bias and may therefore have to wait until more uniform and objective procedures have been discovered. Never-

106

theless, if research into the cognitive processes of classifying is to be done, the kind of experiment described here might suggest some refinements in the descriptive model of classifying in Figure 2 and allow some of the intellectual complexities of the classifying process to be more carefully delineated.

## 5. CLASSIFICATION SYSTEMS AS CONTEXT GRAMMARS

Zsilka[50] suggested that the study of text linguistics should be divided into two domains: text grammar and context grammar. Text grammars include theories of macrocomposition and intratextual linking, such as van Dijk's formulation of how the aboutness of a text is comprehended by readers through a compression of its internal (macro-) propositional content for the purpose of arriving at its highest appropriate macroproposition. Context grammars, according to Zsilka, include theories of the intertextual relationships of a text to other texts and of the extratextual linking of the text to the reality of the moment. We have thus been concerned in this paper to delineate, in Zsilka's terms, the relationship of a text grammar (i.e. van Dijk's theory) to a certain kind of context grammar (i.e. a bibliographic classification system).

The view that a classification system can be seen as a kind of *ad hoc* context grammar was expressed by Shreider and Uspensky, who noted that the 'essential content [of research into library classification] could be described as "empirical ontological semantics"' because it endeavours to 'solve the . . . task of describing some substantial properties of the structure . . . [of] "extralinguistic reality"'.[104] Each classification system both encompasses and is informed by the extralinguistic cultural reality of the whole or of that portion of the world of knowledge it seeks to organise. In this way a bibliographic classification system attempts to provide a relatively formal context grammar by means of which document subjects may be meaningfully analysed and expressed. Svenonius discussed bibliographic classifications from this angle when she wrote 'The hierarchies in which vague words are couched in DDC are perspective hierarchies; they serve not so much to define scientifically as to indicate a point of view or method of treatment. . . A perspective hierarchy functions in retrieval by imbuing vague words with structural meanings and signifying these structural meanings by class numbers.'[105] That is, a system of interlinked perspective hierarchies in a bibliographic classification system offers an interpretative context in which the 'vague words' that describe the conceptual range of the name of a particular class in a particular classification system (and that therefore delimit the general aboutness of documents that can be appropriately entered in the class) can be understood. Once classification has taken place, the documents entered in the class are assumed to be intertextually related both to each other and to the classification system itself.

Basically the same point has been made in the terminology of text linguistics by Zsilka: 'The terser, the poorer, the linguistic information is, the more it is required to include in the communicative process also other systems of signs or other types of signs.'[106] That is, if available linguistic information is terse or poor, as it often is in a classification system, some additional informative signs (e.g. in a classification system, perspective hierarchies, a main class system, established citation orders) are needed to allow the classification text to communicate its own larger intentions and aboutness to the user, in this case the classifier.

Thus, linguistic brevity and vagueness in classification texts means that

linguistic information must be supplemented by the classifier from the overall structure and aboutness of the classification text itself. In these terms, the purpose of the Relative Index in DDC, for example, is to indicate the various perspective hierarchies in which an indexed term or concept appears so that, using the system of signs embodied in the discipline-based main class structure of DDC, the classifier may find the correct context (i.e. an appropriate class name and corresponding notation) for a document. In this sense a classification system provides *contexts for the accommodation of texts that will be classified by it and that are*, as we have seen, intertextually related directly to it and through it to each other. It follows, then, that a specific classification system offers a more or less limited, more or less explicit context grammar for the area of knowledge upon which it is charged with imposing an organisation for the purposes of a certain time and place. The assumption that this assertion is true is responsible, for example, for Ranganathan's conviction that a conceptually structured classified catalogue is more helpful to users than an alphabetic one.[107]

If bibliographic classification systems are seen as empirically developed and applied context grammars, the domains of text linguistics and classification theory may overlap more fully than has been pursued in this essay. Some evidence for this possibility is suggested by the fact that linguists trying to develop a lexicon for the purposes of machine manipulation of natural language investigated the literature of classification theory to see how the world of knowledge has been organised for bibliographic retrieval.[108] Thus, although we have here adopted a theory of discourse or text processing for help in defining aboutness analysis, in identifying common assumptions about intertextuality and in explicating the classifying situation, it may also be that research in classification theory and practice contains insights for text linguistics. Classification research may be seen as a special case of the more general research interests of text linguistics; the kind of research proposed here would then be the kind of genre-specific text analysis advocated by some writers.[e.g. 86]

The theoretical and research interests of text linguistics and of cognitive psychology remain pertinent also to the study of bibliographic classification theory and practice. In general, the field of discourse processing treats the whole subjects of coherent unified texts just as bibliographic classification research and practice do. In particular, exploration of the concept of intertextuality offers a useful area of investigation because subject-based information-organising systems are predicated on the assumption that intertextual links between documents exist and can be discovered and that certain kinds of intertextual links are more useful than others in a user's search for bibliographic material. Empirical research into the classifying process may one day help describe and analyse how intertextual links are identified and interpreted by classifiers. Such clues might tighten a theoretical model of bibliographic classifications as a text type, of the classifiying process as a cognitive act, and of the extent to which the mental processes of assigning topics to documents are shared by classifiers and the eventual users of the documents. In this way one may hope to develop subject retrieval systems capable of generating close matches between a document's aboutness and the meaning(s) the document may have for an individual user with a specific information need.

## REFERENCES

1. MacCAFFERTY, M. 'About aboutness', Co-ordinate Indexing Group colloquium held at Aslib, April 18, 1977. *Journal of Informatics, 1*(3), 1977, 121–122. p. 122.
2. FAIRTHORNE, R. A. Content analysis, specification and control. *ARIST, 4,* 1969, 73–109.
3. HUTCHINS, W. J. *Languages of indexing and classification: a linguistic study of structures and functions.* Stevenage: Peregrinus, 1975. p. 116.
4. CLASSIFICATION RESEARCH GROUP. Bulletin 8. *Journal of Documentation, 20*(3), 1964, 146–165.
5. BOYCE, B. Beyond topicality: a two-stage view of relevance and the retrieval process. *Information Processing and Management, 18,* 1982, 105–109.
6. VAN DIJK, T. A. Relevance assignment in discourse comprehension. *Discourse Processes, 2,* 1979, 113–126.
7. VALDÉS, M. J. *and* MILLER, O., eds. *Identity of the literary text.* Toronto: University of Toronto Press, 1985.
8. ROBERTSON, S. E. Between aboutness and meaning. *In:* MacCAFFERTY, M. *and* GRAY, K., eds. *The analysis of meaning: Informatics 5.* London: Aslib, 1979, pp. 202–205. p. 204.
9. COOPER, W. S. A definition of relevance for information retrieval. *Information Storage and Retrieval, 7,* 1971, 19–37.
10. COOPER, W. S. On selecting a measure of retrieval effectiveness: Part I, The 'subjective' philosophy of evaluation. *Journal of the American Society for Information Science, 24*(2), 1973, 87–100.
11. COOPER, W. S. On selecting a measure of retrieval effectiveness: Part II, Implementation of the philosophy. *Journal of the American Society for Information Science, 24*(6), 1973, 413–424.
12. MARON, M. E. On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science, 28,* 1977, 38–43.
13. SWIFT, D. F., WINN, V. *and* BRAMER, D. 'Aboutness' as a strategy for retrieval in the social sciences. *Aslib Proceedings, 30*(5), 1978, 183–187.
14. PEJTERSEN, A. M. A new approach to the classification of fiction. *In:* DAHLBERG, I., ed. *Universal classification.* Frankfurt: INDEKS-Verlag, 1982. v. 1, pp. 207–215.
15. MARON, M. E. *Op. cit.* p. 42.
16. HUTCHINS, W. J. The concept of 'aboutness' in subject indexing. *Aslib Proceedings, 30* (5), 1977, 172–181.
17. JANOŠ, J. Theory of functional sentence perspective and its application for the purposes of automatic extracting. *Information Processing and Management, 15*(1), 1979, 19–25.
18. METCALFE, J. When is a subject not a subject? *In:* RAWSKI, C. H., ed. *Toward a theory of librarianship.* Metuchen, N. J.: Scarecrow, 1973, pp. 303–339.
19. HUTCHINS, W. J. *Languages of indexing and classification: a linguistic study of structures and functions.* Stevenage: Peregrinus, 1975.
20. HUTCHINS, W. J. On the problem of 'aboutness' in document analysis. *Journal of Informatics, 1*(1), 1977, 17–35.
21. SPECTOR, A. M. Some problems in the analysis of multi-segment texts. *In: The analysis of meaning: Informatics 5.* London: Aslib, 1979, pp. 118–122.
22. BELL, C. *and* JONES, K. P. Back-of-the-book indexing: a case for the application of artificial intelligence. *In: The analysis of meaning: Informatics 5.* London: Aslib, 1979, pp. 155–161. p. 156.
23. VAN DIJK, T. A. Complex semantic information processing. *In:* WALKER, D. C., KARLGREN, H. *and* KAY, M., eds. *Natural language in information science.* Stockholm: Skriptor, 1976, (FID 551), pp. 127–163.
24. DE BEAUGRANDE, R. Text and discourse in European research. *Discourse Processes, 3,* 1980, 287–300.
25. DE BEAUGRANDE, R. *Text, discourse and process: toward a multi-disciplinary science of texts.* Norwood, N. J.: Ablex, 1980.

109

26. BROWN, G. and YULE, G. *Discourse analysis.* Cambridge: Cambridge University Press, 1983.
27. MORGAN, J. L. Discourse theory and the independence of sentence grammar. *In:* TANNEN, D., ed. *Analyzing discourse: text and talk.* Washington, D.C.: Georgetown University Press, 1981, pp. 196–204.
28. DERR, R. L. Linguistic meaning and language comprehension. *Information Processing and Management, 19*(6), 1983, 369–380.
29. DE BEAUGRANDE, R. *Text, discourse and process: toward a multi-disciplinary science of texts.* Norwood, N. J.: Ablex, 1980. p. 16.
30. KATZ, J. J. and FODOR, J. A. The structure of a semantic theory. *In:* FODOR, J. A. and KATZ, J. J., eds. *The structure of language: readings in the philosophy of language.* Englewood Cliffs, N. J.: Prentice Hall, 1964, pp. 497–518.
31. AUSTIN, J. L. *How to do things with words.* 2nd ed. Edited by J. O. Urmson and Marina Sbisà. Cambridge, Mass.: Harvard University Press, 1975.
32. BRANSFORD, J. D. and MCCARRELL, N. S. A sketch of the cognitive aspect of comprehension: some thoughts about what it means to comprehend. *In:* JOHNSON-LAIRD, P. N. and WASON, P. C., eds. *Thinking: readings in cognitive science.* Cambridge: Cambridge University Press, 1977, pp. 377–399.
33. BREWER, W. F. and LICHTENSTEIN, E. H. Event schemas, story schemas, and story grammars. *In:* LONG, J. and BADDELEY, A., eds. *Attention and performance IX.* Hillsdale, N.J.: Lawrence Erlbaum, 1981, pp. 363–379.
34. VAN DIJK, T. A. and KINTSCH, W. Cognitive psychology and discourse: recalling and summarizing stories. *In:* DRESSLER, W. U., ed. *Current trends in text linguistics.* Berlin: Walter de Gruyter, 1978, pp. 61–80.
35. VAN DIJK, T. A. *Macrostructures: an interdisciplinary study of global structures in discourse, interaction, and cognition.* Hillsdale, N.J.: Lawrence Erlbaum, 1980.
36. VAN DIJK, T. A. *Text and context; explorations in the semantics and pragmatics of discourse.* London: Longman, 1980. p. 131. (original capitalisation)
37. KINTSCH, W. and VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review, 35*(5), 1978, 363–394. p. 363.
38. NAJARIAN, S. E. Organizational factors in human memory: implications for library organization and access systems. *Library Quarterly, 5*(3), 1981, 269–291.
39. VAN DIJK, T. A. *Text and context: explorations in the semantics and pragmatics of discourse.* London: Longman, 1980, pp. 133–134. (original capitalisation)
40. VAN DIJK, T. A. Complex semantic information processing. *In:* WALKER, D. C., KARLGREN, H. and KAY, M., eds. *Natural language in information science.* Stockholm: Skriptor, 1976, (FID 551), pp. 127–163. p. 137.
41. KINTSCH, W. and VAN DIJK, T. A. *Op. cit.* p. 374.
42. VALDÉS, M. J. and MILLER, O., eds. *Op. cit.* p. vii.
43. KIERAS, D. E. Initial mention as a signal to thematic content in technical passages. *Memory and Cognition, 8*(4), 1980, 345–353.
44. KIERAS, D. E. Topicalization effects in cued recall of technical prose. *Memory and Cognition, 9*(6), 1981, 541–549.
45. VAN DIJK, T. A. Recalling and summarizing complex discourse. *In:* BURGHARDT, W. and HOLKER, K., eds. *Text processing: papers in text analysis and text description.* Berlin: Walter de Gruyter, 1979, pp. 49–118.
46. VAN DIJK, T. A. Episodes as units of discourse analysis. *In:* TANNEN, D., ed. *Analyzing discourse: text and talk.* Washington, D.C.: Georgetown University Press, 1981, pp. 177–195.
47. RANGANATHAN, S. R. *The five laws of library science.* 2nd ed. Bombay: Asia Publishing, 1963.
48. RANGANATHAN, S. R. *Prolegomena to library classification.* 3rd ed. N.Y.: Asia Publishing, 1967.
49. VAN DIJK, T. A. *Macrostructures: an interdisciplinary study of global structures in discourse, interaction, and cognition.* Hillsdale, N.J.: Lawrence Erlbaum, 1980. p.v.

50. ZSILKA, T. Communicative relationships of the text. *In:* PETÖFI, J. S., *ed. Text vs sentence.* Hamburg: Helmut Buske, 1981, v. 3, pp. 231–240. p. 233.
51. VAN DIJK, T. A. *and* KINTSCH, W. *Op. cit.* p. 72.
52. DE BEAUGRANDE, R. *Text, discourse and process: toward a multi-disciplinary science of texts.* Norwood, N.J.: Ablex, 1980. p. 242.
53. *Ibid.* p. 20.
54. FAIRTHORNE, R. A. Temporal structure in bibliographic classification. *In:* WOJCIE-CHOWSKI, J. A., *ed. Conceptual basis of the classification of knowledge: proceedings of the Ottawa Conference,* 1971. Munich: Verlag Dokumentation, 1974, pp. 404–412.
55. STEINER, G. *After Babel: aspects of language and translation.* London: Oxford University Press, 1975.
56. STEINER, G. *On difficulty and other essays.* Oxford: Oxford University Press, 1978.
57. FAIRTHORNE, R. A. *Towards information retrieval.* London: Butterworths, 1961. p. 182.
58. HARRIS, B. *and* HOFMANN, T. R. Regularized abstracting. *In: Informatics 1.* London: Aslib, 1974, pp. 95–108.
59. INFOTERM. *International cooperation in terminology.* Munich: Verlag Dokumentation, 1976.
60. DE BEAUGRANDE, R. *Text, discourse and process: toward a multi-disciplinary science of texts.* Norwood, N.J.: Ablex, 1980. pp. 19–20.
61. BLAIR, D. C. *Pragmatic aspects of inquiry.* Ph.D. Dissertation, University of California at Berkeley, 1981. (*Dissertation Abstracts* no. 8200026) p. 5.
62. *Ibid.* p. 153.
63. CLIFTON, C., SORCE, P. *and* CRUSE, D. The translation effect in memory search. *Cognitive Psychology, 9,* 1977, 1–30.
64. DAHLBERG, I. The terminology of subject fields. *In:* INFOTERM. *International cooperation in terminology.* Munich: Verlag Dokumentation, 1976, pp. 61–71. p. 62.
65. VAN DIJK, T. A. Complex semantic information processing. *In:* WALKER, D. C., KARLGREN, H. *and* KAY, M., *eds. Natural language in information science.* Stockholm: Skriptor, 1976, (FID 551), pp. 127–163. p. 140.
66. *Ibid.* p. 141.
67. BLISS, H. E. *The organization of knowledge and the system of the sciences.* New York: Henry Holt, 1929. p. 133.
68. LEE, J. M. E. Wyndham Hulme: a reconsideration. *In:* RAYWARD, W. B., *ed. The variety of librarianship.* Sydney: Library Association of Australia, 1976, pp. 101–113.
69. BULICK, S. *Structure and subject interaction: toward a sociology of knowledge in the social sciences.* New York: Marcel Dekker, 1982.
70. BLISS, H. E. *A bibliographic classification.* New York: H. W. Wilson, 1940–53.
71. BLISS, H. E. *The organization of knowledge and the system of the sciences.* New York: Henry Holt, 1929. p. 209.
72. RANGANATHAN, S. R. *The Colon Classification.* New Brunswick, N.J.: Rutgers University Press, 1965. p. 22.
73. DAHLBERG, I. ICC – Information Coding Classification – principles, structure and application possibilities. *International Classification, 9(2),* 1982, 87–93.
74. AUSTIN, D. Citation order and linguistic structure. *In:* RAYWARD, W. B., *ed. The variety of librarianship.* Sydney: Library Association of Australia, 1976, pp. 19–46.
75. DAHLBERG, I., *ed. Universal classification: proceedings of the 4th International Study Conference on Classification Research.* 3v. Frankfurt: INDEKS-Verlag, 1982.
76. GRAZIANO, E. E. *Language-operational-gestalt awareness.* Tempe, Ariz.: Association for Library Automation Research Communications, 1975. p. 351.
77. BRITAIN, I. C. S. Classification and culture. *Australian Academic and Research Libraries, 6,* 1975, 31–44.
78. LINDEN, R. Some thoughts on the Dewey Decimal Classification. *Library Resources & Technical Services, 26,* 1982, 183–187.
79. HULME, E. W. *Statistical bibliography in relation to the growth of modern civilization.* London: Butler and Tanner, 1923.

111

80. HULME, E. W. Principles of book classification. *Library Association Record, 13,* 1911, 354–358; 389–394; 444–449; *14,* 1912, 34–46; 174–181.

81. PRITCHARD, A. Statistical bibliography or bibliometrics? *Journal of Documentation, 25,* 1969, 348–349.

82. DE GROLIER, E. Classifications as cultural artefacts. *In:* DAHLBERG, I., ed. *Universal classification.* Frankfurt: INDEKS-Verlag, 1982, v.2, pp. 19–34. p. 33.

83. RANGANATHAN, S. R. *The Colon Classification.* New Brunswick, N.J.: Rutgers University Press, 1965. p. 140.

84. VAN DIJK, T. A. New developments and problems in text linguistics. *In:* PETÖFI, J. S., ed. *Text vs sentence.* Hamburg: Helmut Buske, 1981, v.2, pp. 509–523. p. 521.

85. KINTSCH, W. *and* VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review, 35*(5), 1978, 363–394. p. 373.

86. BLACK, J. B. The effects of reading purpose on memory for text. *In:* LONG, J. *and* BADDELEY, A., eds. *Attention and performance IX.* Hillsdale, N.J.: Lawrence Erlbaum, 1981, pp. 347–361. p. 358.

87. MAYORDOMO, T. A. On text linguistic theory. *In:* PETÖFI, J. S., ed. *Text vs sentence.* Hamburg: Helmut Buske, 1981, v.3, pp. 1–15.

88. BÁNRÉTI, Z. The topic of texts and the interpretation of texts. *In:* PETÖFI, J. S., ed. *Text vs sentence.* Hamburg: Helmut Buske, 1981, v.3, pp. 43–57.

89. LANCASTER, F. W. Information science. *In:* WALKER, D. C., KARLGREN, H. *and* KAY, M., eds. *Natural language in information science.* Stockholm: Skriptor, 1976, (FID 551), pp. 19–43. p. 22.

90. VAN DIJK, T. A. Relevance assignment in discourse comprehension. *Discourse Processes, 2,* 1979, 113–126. p. 118.

91. *Ibid.* p. 119.

92. ROLLING, L. Indexing consistency, quality and efficiency. *Information Processing and Management, 17,* 1981, 69–76.

93. MARKEY, K. Interindexer consistency tests: a literature review and report of a test of consistency in indexing visual materials. *Library and Information Science Research, 6,* 1984, 155–177.

94. MOELLER, B. A. Subject analysis in the library: a comparative study. *International Classification, 8*(1), 1981, 23–37. p. 23.

95. CLEVERDON, C. W. *Aslib Cranfield Research Project: report on the testing and analysis of an investigation into the comparative efficiency of indexing systems.* Cranfield, 1962.

96. CAMPBELL, D. T. *and* STANLEY, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1963.

97. OLTON, R. M. Experimental studies of incubation: searching for the elusive. *Journal of Creative Behavior, 13*(1), 1979, 9–22.

98. RANGANATHAN, S. R. *Library classification: fundamentals and procedures.* Madras: Madras Library Association, 1944.

99. REDER, L. M. *and* ANDERSON, J. R. A comparison of texts and their summaries: memorial consequences. *Journal of Verbal Learning and Verbal Behavior, 19,* 1980, 121–134.

100. REDER, L. M. *and* ANDERSON, J. R. Effects of spacing and embellishment on memory for the main points of a text. *Memory and Cognition, 10*(2), 1982, 97–102.

101. TELL, B. V. Document representation and indexer consistency. *In: Cooperating information societies: Proceedings of the American Society for Information Science.* Westport, Conn.: Greenwood Press, 1969, v. 6, pp. 285–292.

102. MEYER, B. J. F. *The organization of prose and its effects on memory.* Amsterdam: North-Holland, 1975.

103. RUMELHART, D. E. Understanding and summarizing brief stories *In:* LABERGE, D. *and* SAMUELS, S. J., eds. *Basic processes in reading: perception and comprehension.* Hillsdale, N.J.: Lawrence Erlbaum, pp. 265–303.

104. SHREIDER, YU. A. *and* USPENSKY, V. A. Semantic aspects of informatics. *In: Research on the theoretical basis of information.* Moscow: FID, 1975, (FID 530), pp. 152–169. p. 169.

105. SVENONIUS, E. Indexical contents. *In:* DAHLBERG, I., *ed. Universal classification.* Frankfurt: INDEKS-Verlag, 1982, v.3, pp. 125–138. p. 133.
106. ZSILKA, T. Communicative relationships of the text. *In:* PETÖFI, J. S., *ed. Text vs sentence.* Hamburg: Helmut Buske, 1981, v.3, pp. 231–240. p. 236.
107. RANGANATHAN, S. R. *Classified catalogue code.* 5th ed. Bombay: Asia Publishing, 1964.
108. NEUMANN, R. Thesauri und Klassifikation von Wissen. *In:* PETÖFI, J. *and* BREDEMEIER, J., *eds. Das Lexikon in der Grammatik - die Grammatik im Lexikon.* Hamburg: Helmut Buske, 1977, pp. 163–277.