PERSPECTIVE PAPER:

COMPLEX SEMANTIC INFORMATION PROCESSING

Teun A. van Dijk
University of Amsterdam

Introduction - Aims - Problems

In this paper some recent results from linguistics, cognitive psychology and artificial intelligence will be applied in the formulation of some major problems of information science.* These results pertain to theories about the structure and processing of discourse and of complex semantic information in particular. It is argued that grammatical and other models of discourse constitute a necessary basis for a sound theoretical explication of such terms as 'content of a text (or document)', 'abstract' or 'summary', 'key-word', 'paraphrase', 'theme' or 'topic', etc., as they are used in information science. In this respect the current automatic or non-automatic approaches to the analysis and processing of texts in documentation are found to be inadequate, whatever their practical merits may be.

The background for the proposals in this paper is the development, in linguistics, of so-called *text grammars*** and of other theories of discourse structure and discourse processing in neighbouring disciplines such as cognitive psychology, artificial intelligence, poetics, anthropology, etc. These various, in part inter-disciplinary, approaches will be captured under the label of *discourse studies*. Attempts towards the elaboration of some elements of text grammars began in the late sixties, particularly in both Germanies, and were paralleled by other work paying increasing attention to pragmatic, social and cultural *contexts* of language and language use. Both directions of research are to be seen as reactions against some major tenets in the generative-transformational ("Chomskyan") paradigm in linguistics, e.g., the study of the structure of sentences in isolation. It has been shown that the syntactic and semantic structure and interpretation of sentences should be studied relative to those of other sentences in the discourse. Moreover, it has been argued that discourses may have specific overall structures, which cannot be formulated in terms of sentence structure alone, e.g. semantic *macrostructures*.

----

*It should be emphasized that only a few aspects of the possible applications of linguistics are dealt with in this paper. Secondly, the paper is strictly informal: no logical definitions of the various structures and rules are given. Below reference will be made to other work on which this paper is based. Finally, our knowledge of information science is limited. It may well be that recent developments in this field would make some of our critical remarks about traditional 'content analysis' superfluous.

**Work on text grammar is too voluminous to be fully considered here. Dressler & Schmidt's bibliography (1973) lists hundreds of titles, although including work on related topics. For references, introductions, and surveys in English, see, e.g., van Dijk (1972a, 1976b), Petöfi and Rieser (1973). Buske Verlag in Hamburg publishes a series of preprints and books on text linguistics. Van Dijk & Petöfi (1976) bring together a collection of studies in which various authors in text linguistics apply their methods and proposals in the analysis of a single text (a fable by Thurber). Their book is the first volume of a new series on text linguistics by De Gruyter (New York-Berlin), in which also the first larger English survey (Dressler, 1977) will be published.

Teun A. van Dijk

In poetics and anthropology it was shown that the analysis of different kinds of discourse, e.g., of narratives, must be given in terms of units, categories, and rules based on the semantic macro-structures of the discourse.* This insight holds for the analysis of specific types of discourse in general (arguments, news reports, conversations, scientific articles, propaganda, etc.).

In sociology, socio-linguistics and pragmatics the basic idea has been further developed that the structure of language systems and utterances determines and is determined by the functions of language in communication and social interaction. More in particular it has been found that everyday conversation is bound by rather strict conventional rules,** and that structures of discourse run parallel with structures of speech act sequences.***

In cognitive psychology,**** finally, it has been pointed out that verbal processing is basically semantic, and that the production and understanding of complex discourse is not sentential but co-determined by macro-structures and mental representations of conventional knowledge structures. 'frames' as they are called in recent artificial intelligence approaches to higher information processing and language understanding.*****

The details of these various directions of research constituting the background of this paper cannot possibly be spelled out here. We will have to focus attention on those assumptions and results which can profitably be brought to bear in the domain of information storage and retrieval problems of documentation. More in particular, the highly intricate issue will be attacked concerning the assessment of the semantic 'content' of texts, not only of their individual sentences but also of the text as a whole. In other words. we will address problems in a domain which may be called that of *complex semantic information processing*. The theoretical apparatus to be developed not only provides insight into *natural* (cognitive) information processing, but also into some crucial tasks of *artificial* information processing, e.g., in documentation. The two basic concepts involved in solving problems of complex information processing are *organization* and *reduction*. It will be shown that the theory of semantic macro-structures provides the basis for an explanation of how large and complex amounts of information are organized and reduced. This theory, in particular, allows for the first time an explicit account of the structure and formation of *abstracts* of texts.

----

*See for references (van Dijk, 1972a, 1975a,b).

**This work has been done in an area of research called 'ethnomethodology'. Recent readers containing work on conversation are Sudnow (1972), Cicourel (1973), Turner (1973). Similarly, under the label of the 'ethnography of speaking', anthropologists have recently been studying the various kinds of culturally typical discourses (see Bauman & Sherzer, 1974).

***The last part of van Dijk (1976b) is dedicated to the systematic interdependencies of discourse structure and the structure of speech acts and speech act sequences.

****Developments in cognitive and experimental psychology on the processing of discourse run parallel with those in (text) linguistics. See Freedle and Carroll (1972), Kintsch (1974), Meyer (1975), van Dijk & Kintsch (1975, 1976), van Dijk (1975a).

*****Closely related to the mentioned developments in psychology, artificial intelligence has also had increasing interest in 'higher' information processing in recent years, e.g., discourse (see Charniak, 1972). For the notion of 'frame', see especially Minsky (1975), Bobrow and Collins (1975), Norman & Rumelhart (1975). For the relations between frames and macro-structures, see our contribution to the 1976 Carnegie-Mellon Cognition Symposium on Discourse Comprehension (van Dijk, 1976c).

Perspective Paper: Complex Semantic Information Processing

The Structure of Discourse

We start from the assumption that both natural and artificial discourse processing is in part based on the structures assigned to the discourse. Such structures are made explicit in the structural descriptions provided by a theory of discourse, e.g., by a text grammar. These descriptions are abstract, whereas the text descriptions used in information storage and retrieval are not; these are also discourses of a specific kind, and must be based on abstract description, on the one hand, and on the 'pragmatic' properties of the use of information in retrieval processes of various kinds, on the other hand. In an account of the abstract structures of discourse, we will further make the assumption that information processing, especially in documentation, is primarily *semantic*. This means that morphonological, syntactic, stylistic, and rhetorical properties of discourse will be ignored in this paper. It is understood, of course, that the 'content', i.e., the semantic representation, of a discourse is expressed by, or accessible via, these 'surface structures' of the discourse. Any full-fledged system of information processing will therefore have to include a surface grammar of some kind, relating these structures to the 'underlying' semantic representation. Since only partial surface grammars of this type are available at the moment, any workable information processing system is necessarily incomplete or ad hoc. Nevertheless, the analysis of semantic information from discourse is relevant not only for semantic theory. Surface structure analysis grammars or programs must be developed with an eye on the semantic representations; as is the case in logic, an explicit syntax, for instance, will have to reflect the units, categories, and rules of semantic interpretation. Current generative-transformational syntax does not satisfy this condition, whereas recent proposals in categorial grammars are too limited and complex to allow direct practical application, although they are more adequate from a semantic and logical point of view.*

Before we tackle the problem of the semantic structure of discourse, some terminological issues must be addressed. First of all, as was indicated above, the notion of *information* will be taken in its semantic sense, pertaining to the *meaning* and *reference structure* of a verbal utterance. Meaning and/or reference is assigned to an expression of a natural or artificial language by the 'real' or formal process of *interpretation*. That is, we neglect for a moment the pragmatic properties of information, as they are defined in terms of the knowledge or acts (e.g., assertions, questions) of language users in communicative contexts. Nor will we speak of the various kinds of 'structural' information a discourse may have, e.g., of a syntactic or stylistic nature, whatever the rhetorical/pragmatic effects of such information in communication and interaction.

Secondly, we would like to distinguish between different theoretical notions of *discourse* and *text*. A discourse will be taken as an empirical, cognitive and social, verbal unit satisfying a number of specific conditions (continuity in time, discreteness, etc.). A discourse is physically manifested by *verbal utterances* (tokens), which we can hear or see, but which we conventionally understand 'as' discourses of a natural language. As will be made clear below, only those utterances may conventionally count as (i.e., be produced, understood, accepted) discourses of a language which have underlying *text* structure. That is, a text is an abstract theoretical construct of a grammar (or other theory of discourse), making explicit the structure of a discourse. The everyday, intuitive use of the term 'text' in English (often meaning 'written/printed piece of discourse') will therefore be ignored

----

*For recent work on these developments in logical grammar, categorial syntax, and formal semantics, see Cresswell (1973), Montague (1974), Hintikka, Moravcsik, and Suppes (1973), and Keenan (1975). In this paper several concepts from logical semantics, such as 'interpretation', 'possible world', 'domain of individuals', are used. We refer to the work mentioned above for further explication of these terms, and to van Dijk (1976b) for an introduction and application to the semantics of discourse.

here. Finally, the term *document*, as usual, will denote a concrete object 'containing' or 'carrying' one or more discourses, and having a specific processing function (storage, analysis, abstracting, retrieval, etc.). Thus, we can *read* a document, but only *understand* one of its discourses, whereas the understanding of a discourse takes place by the cognitive or theoretical assignment of text structure to that discourse. Other terminological introductions and distinctions will be made below.

Above, two different kinds of *semantics* have been mentioned, namely, a semantics accounting for the *meaning* of a discourse, and a semantics accounting for the *reference* of a discourse. The first kind of semantics has been predominant in linguistics, the second in logic and, partly, in the philosophy of language.* A serious account of discourse structure, however, needs both kinds of semantics, which, in fact, are intimately related. On the one hand we may only know what words, phrases or sentences 'denote' or 'refer to' if we know what they 'mean'. On the other hand, the conventional and conceptual meaning of expressions can be made explicit only in terms of their potential ability to 'determine' reference with respect to individual objects, properties, relations, facts (truth values) in certain possible worlds. Thus, as we will show in a moment, the relations involved in the fundamental notion of discourse coherence are to be formulated in terms of meaning semantics and in terms of reference semantics, determining, for example, that two expressions, although different in meaning, may refer to the same individual. The task of semantics, in general, is to give recursive conditions determining the meaning or reference of sentences (or propositions) on the basis of the meaning or reference of their constituents.

Similarly, an explicit *textual* semantics must formulate the meaning or reference of a discourse in terms of the meaning or reference of its constituent sentences. In order to denote the meaning of an 'atomic' sentence, we use the term *proposition*. Several propositions may be *expressed* by one (composite) sentence. A proposition will be taken to have the usual structure consisting of an n-place predicate, n arguments, argument labels (cases), and various kinds of preceding operators and quantifiers. Propositions, just like their constituents, are taken as abstract *concepts*, i.e. as *functions* taking values (extensions) in some possible world. Instead of the usual truth values we shall use the notion of *fact* for the extension of a proposition. Conversely, a fact is an element or property of a possible world as *represented* by a proposition.

The semantic structure of discourse will be characterized at two different, though connected, levels, namely, the level of *micro-structures* and the level of *macro-structures*. By the term 'micro-structure' -- which is only used for practical reasons, we mean the structure of propositions and the linear structure of sequences of propositions in a discourse. By macro-structure, we mean a level of description pertaining to the semantic structure of discourses as a whole or parts of discourses taken as wholes (units). Before we make this distinction explicit, an intuitive example illustrating it may be given. The proposition sequence as expressed by

> *(1) Peter married Laura.   She became very unhappy.*

may be representing two particular facts and some relations between them, micro-structurally, but at the macro-structural level the propositions may be used to represent a very complex series of facts, as in a full story about Peter meeting, marrying, etc., Laura.

----
*See the references in the previous footnote.

Perspective Paper: Complex Semantic Information Processing

In other words: a micro-structural analysis provides the *local* structure of proposition sequences, whereas the macro-structural analysis yields a *global* structure of the proposition sequence. The usual linguistic and logical semantics has been limited to this kind of 'micro-semantics'. Informal approaches to macro-semantics come from poetics, anthropology, etc. The 'plot' of a story, for example, may be formulated in terms of macro-structural units, such that each unit may consist of a whole set of propositions. A similar global distinction exists between the premise(s) and the conclusion of an argument or proof, or between the different conventional parts of a scientific paper, a novel or a drama.

In order to satisfy the basic requirement of semantics that 'wholes' must be defined in terms of their 'parts', macro-structures must be derived from micro-structures. In order to be able to do this, macro-structures are also described in terms of propositions, because they are also meaning/reference structures, although at another level. The mappings relating the sequence of propositions underlying the actual sentences of the discourse with the sequence at some macro-level will be called *macro-rules*. The formulation of empirically adequate macro-rules is one of the major problems of actual text grammars, and will be attempted below.

There are various linguistic and cognitive arguments which have led to a distinction between a linear, local level of meaning/reference and levels of global meaning/reference of a discourse. Not only do we want to account for something like the meaning of a discourse 'as a whole', but at the same time the notion of macro-structure is important in the characterization of linear (local) coherence conditions of composite sentences and sequences. Cognitively, macro-structures have been shown to be indispensable in order to explain our ability to produce, understand, store and (re-)organize, summarize, recall, and infer from highly complex information structures, such as discourses. We return to these linguistic and psychological facts below. The semantic micro-structure, to begin with, is described as an ordered n-tuple of propositions.* However, not every sequence of propositions is interpretable, and uninterpretable proposition sequences determine the non-acceptability of the discourses expressing them. Hence, there are cooccurrence conditions for proposition sequences. A sequence satisfying the conditions, at this level, will be called *linearly coherent.*

A first set of coherence conditions pertains to pairwise connections between propositions. These *connections* may be expressed by various types of natural *connectives*, e.g., *and, but, although, or, yet, because, for, so, if...then,* etc., relating clauses in composite sentences and sentences in sequences. In order to be able to connect two simple or composite propositions, a certain number of *connection conditions* must be satisfied. These conditions, first of all are to be given in terms of *reference*. Globally speaking, then, two propositions are connected if their 'referents', i.e., the facts they denote in some possible world or course of events, are related. These relations are of various types, differing in 'strictness'; they range from possibility, via probability, to different kinds of necessity. Facts may just cooccur in some situation (world-time point or period), or they may allow each other to occur or necessitate each other, in at least one, in most, or in all possible

----

*The notion of 'connection' or 'relevance' not only has been treated in text grammars (see van Dijk, 1976b), but also occurs in recent development in logics, especially so-called 'relevance (or entailment) logics', where a number of intuitively unsatisfactory theorems of classical logic (e.g., ex falso sequet quod libet) are no longer valid. That is, it is somehow assumed, especially when modality, e.g., entailment, is involved, that propositions which are connected by logical connectives or related by logical inference rules, should be 'relevant to' each other, either in meaning or in truth conditions, or both. The technical proposals cannot be gone into here. For brief surveys and for the differences between logical and natural connectives, see van Dijk (1976a, 1974).

courses of events. These distinctions define the different kinds of connection as expressed by conjunctions, disjunctions, concessives, causals and implications. Other differences between connectives of the same kind (e.g. *but/although, because/for*) are to be accounted for in presuppositional and (other) pragmatic terms.

These relations between facts in possible courses of events often involve relations between individuals or properties/relations. Thus, in sequence (1) above, there is a *referential identity* relation: the referent of the expression *Laura* in the first sentence is the same as that of the expression *she* in the second sentence. It is in this sense that we say that the *interpretation* of proposition sequences is *relative*: the second sentence (proposition) must be interpreted relative to the interpretation of the first sentence (proposition), because the domains of individuals involved intersect. Note, however, that referential identity, as such, is neither sufficient nor necessary:

> *(2) Peter married Laura. She has a sister called Mary.*

> *(3) Peter married Laura today. The church was full of friends.*

In (2) there is referential identity but, normally, the facts denoted are not directly related. In (3) there is no overt referential identity, but the two respective facts are related by reason/cause.

The connection conditions are themselves *relative*. First of all, what is a connection for *A* need not always be a connection for *B*, e.g., when reasons are involved. These differences, however, are pragmatic and will be abstracted from in the semantics, also because we adopt the assumption that connections are also conventional, i.e., of a more general nature, based on common knowledge of the postulates governing the actual or alternative possible worlds. Secondly, however, connection is relative to what will be called a *topic of discourse*, or more in general a *topic of conversation*. Connected propositions may denote related facts, but these may belong to completely different topics, intuitively speaking. In other words, when we express a number of propositions in a sentence or sequence, these propositions must somehow belong to the same 'range' of semantic space. Clearly, we here meet an *intensional* condition, pertaining to the conceptual meanings involved. Thus marriage and (un-)happiness may characterize the same range of 'human' or 'action' concepts, whereas marriage and liquidity, or having a sister and reading a book do not. More generally, not only are conceptual meaning relations involved in topical identity, but also other properties of conventional knowledge, namely, *frames*: we know, conventionally, that a marriage may take place in a church, which makes sequences like (3) connected.

A topic of conversation will be formally represented as an (ordered) set of propositions. In particular, it will be argued below that such a topic should be defined at the level of macro-structures. The same holds for the conditions determining *topic changes*.

There are also coherence relations in proposition sequences between non-connected (sequences of) propositions. The fact, for example, that the individual denoted by expressions in a whole sub-sequence of propositions may remain the same (thus constituting a so-called *discourse referent*) -- even if the facts denoted by these propositions are not pairwise related -- is one of the possible conditions making sequences coherent. More specifically, each individual introduced in the domain of interpretation ('universe of discourse') must be related to an individual already introduced. Similarly, the possible worlds in which the denoted facts occur must be identical or related (by accessibility, for instance). Finally, the more general condition pertaining to the topic of

conversation also holds here: the individuals and/or their properties must belong to the same semantic range or the same system of interrelated frames. Examples will be given below.

Summarizing the various conditions of linear discourse coherence, it might be said that each proposition of the sequence must be interpreted relative to the interpretation of at least one other proposition, and relative to a common topic of discourse (or alternative topic, initiatable, i.e., accessible, from a given topic). Note that these relative interpretations involve both referential and meaning properties of sentence sequences.

There is another point which should be made about the description of linear coherence in discourse, and which has important implications for information science. We have adopted the condition that each proposition of a sequence underlying a discourse should be interpreted relative to another proposition of the sequence. This condition, however, needs further qualification. It is very often the case in natural discourse, as distinguished from formal discourse. e.g., in proofs, that much of the information remains *implicit*. That is, in order to be able to postulated coherence, propositions may be 'present' without being (fully) expressed in the surface structure of the discourse. In order to be able to explain the definite article in a sequence like

> *(4) (...) We ordered two pizzas. The waitress, however, brought us two spaghettis.*

we must assume that a waitress is introduced by an implicit proposition, 'implied' by the information that we are in a restaurant. These implicit propositions are present in the theoretical sequence of the discourse (the so-called *explicit text base*) due to (i) the *meaning postulates* of the language (once we have the information that John is a bachelor, we also know that he is a male adult and that he is not married), (ii) the *conventional knowledge (frames)* implied by one or more propositions (we know that in a restaurant we can order food, and that our order is taken and carried out by a waiter or waitress). Since these weakly or strongly implied propositions are generally known, a pragmatic rule allows them not to be expressed explicitly. To make sure that the explicit text base does not become too rich with (infinite) sets or entailed propositions, we stipulate that only those propositions are part of the text base which are necessary conditions for the correct interpretation of other (expressed) propositions of the sequence, as was the case for the interpretation of the definite noun phrase in the second sentence of (4). Below, we will show that the macro-information of a discourse also need not be, and is not generally, expressed in the discourse itself by direct expression in surface structure.

The implication for information science is that a discourse may well contain information which, as such, is not expressed in surface structure, so that a pure surface analysis or a semantic description of the expressed propositions may yield an incomplete characterization of the content of a document. Hence, any theoretically adequate account of the content of a discourse must be based on the explicit text base as defined above.

More generally, it should be emphasized that *information* as defined is essentially *propositional*. Hence an internal propositional structure must be present, not, for instance, a simple predicate or argument (as expressed by some surface noun, pronoun or verb). In order to be able to assess this propositional structure, at least a partial syntactic analysis of the surface structure of the discourse is necessary. It follows that any word-based analysis of information is inadequate.[*]

----

[*]For references to various kinds of 'word-based' document analyses, see Sparck Jones and Kay (1973). For work on content analysis see, e.g., Gerbner et al (1969) and Holsti (1969), especially in the social sciences. We are ignorant about possible new developments in the latter field.

Teun A. van Dijk

Macro-Structures of Discourse

Linear connection, and coherence of discourse in general, is, as we saw, also determined by a more global 'topic of conversation'. This notion will be explained in terms of *macro-structures* of discourse.*

Intuitively, then, a macro-structure of a discourse represents its overall meaning, and the global meaning constituents forming this overall meaning. Formally, a macro-structure is also a sequence of propositions, and is acquired by mappings, so-called *macro-rules*, from the full text base of the discourse.

Macro-structures define the *overall coherence* of a discourse. In other words, a sequence of propositions (as part of a text structure) may underlie an acceptable discourse only if this sequence is macro-structurally coherent. It is not difficult to illustrate this point. A sequence may be linearly coherent without being globally coherent:

> *(5) I bought a book in the bookshop this morning. The bookshop has been rented by a French woman. French women are usually well-dressed. My sister dresses well, too . . .*

Intuitively, we know that such sequences do not have a 'point' or 'theme' or unified 'topic'. We assume that we can make this intuition and these notions explicit by requiring any discourse to have a macro-structure. Such a macro-structure in a sense is a global constraint on the individual propositions of the sequence: it defines the intensional ranges for the different concepts used, and defines the worlds and types of facts referred to. In particular, it allows the individual proposition to have a specific function in a larger structural whole, as we will see below.

The cognitive importance of macro-structures has been assessed in several recent experiments about the comprehension, recall and summarization of discourse, e.g., of stories.** Sequences without a macro-structure, just like sequences of words without a sentence structure, are much less well recalled. The further general assumption is that subjects are unable to store the individual sentences and propositions of a discourse. Nevertheless, they are able to give a summary of the discourse as a whole. This fact can only be explained if they *organize* and *reduce* the large amount of propositional information during input. In fact, they apply macro-rules such that shorter macro-structures are obtained. These structures can be used, later, as recall cues in the search for detailed information. After longer delays, however, not much more than the macro-propositions are available in free recall. Such delayed recall protocols bear a striking similarity to immediate *summaries*. Hence, it should be assumed that a summary of a discourse is a discourse expressing the/a macro-structure of the original discourse. Controlled summarizing, then, is based on the same rules and processes as noncontrolled forgetting. Below, we will apply these results (of which no further experimental or theoretical details can be given here) in a model for artificial information processing.

----

*We have been using the term 'macro-structure' since about 1968 in various papers and books in poetics, first, then in linguistics, a first survey being given in van Dijk (1972a). We have taken up the originally very informal and unprecise ideas again in 1974, using insights from formal semantics and developments in cognitive psychology. The actual state of macro-structure theory is formulated in van Dijk (1976c). Other uses of the term have been made, see e.g., Ballmer (1975). Our own ideas on this topic have considerably changed since the late sixties, especially concerning the possible role of the grammar (semantics) in an account of macro-structures.

**See the references on cognitive and experimental psychology cited earlier. Our own experiments are reported in van Dijk (1975a) and Kintsch and van Dijk (1975, 1976). For a recent survey of other work see Thorndyke (1975).

It should be added that macro-structures not only determine the execution of complex linguistic tasks, but also are necessary for the planning, execution, control, and comprehension of any kind of highly complex behavior, e.g., in the visual interpretation of complex scenes, in the solutions of problems, and in (inter-)action.

In order to make macro-structures explicit, we assume at least four *macro-rules*, namely:

I. DELETION

II. GENERALIZATION

III. SELECTION

IV. CONSTRUCTION or INTEGRATION

These rules make explicit the usual ways in which we intuitively organize and reduce our information, either during input or within our memory (knowledge) systems. The first rule, DELETION, allows that all propositions may be deleted from an explicit text base, which are not presuppositions (conditions) for the interpretation of subsequent propositions of the sequence. Thus in the sequence:

(6) Mary was playing with a red ball. The ball smashed a window.

the proposition *red(a)*, where *a* is a ball, may be deleted, because it is *globally irrelevant* for the interpretation of the discourse, whatever the local relevance may be. We see that the macro-rules also function as an explicit device to define what information is *relevant* or *important* in a discourse, semantically speaking at least (we neglect the pragmatic implications of these notions, having to do with conditions and consequences of [speech] interaction of the speaker and the hearer).

Whereas the first rule is intended to delete irrelevant 'attributes', the second rule, GENERALIZATION, abstracts from inherent properties of individuals. It allows propositions with a concept *a* being replaced by propositions with a superconcept of *a*. The organizational and reductional effect of this rule in particular lies in the fact that a whole sequence of propositions with concepts *a, b, c, ...* , may be reduced to a single proposition with the superconcept of *a, b, c.*:

(7) Mary was playing with her doll house. She played with her blocks, then with her racing cars (. . .).

This sequence may be substituted by a proposition underlying

(8) Mary was playing with her toys.

In both rules, the information of the text base becomes formally irrecoverable by the application of the rules. If we would only have the macro-propositions, we could only construct an infinite set of possible underlying text bases (with the same macro-structure).

In the following two rules, this is not the case. The information left out here is partly re-constructible. The SELECTION rule is applied on those sequences in which some proposition weakly or strongly implies the other propositions, either by meaning postulate or by frame information. The implied propositions may then be deleted:

135

Teun A. van Dijk

(9) That day we went to Paris. In the morning we took a cab to the station, there we took the train (. . .).

From this sequence we need just keep the proposition underlying the first sentence, since the other proposition are weakly implied by it: if we travel to another town, we must take some means of transportation, like a train or plane, and execute preparatory actions to accomplish the main action. In general what may be deleted in such sequences are

(10) --normal conditions and presuppositions (e.g., preparatory actions)

--normal component actions, either optional or necessary

--normal consequences, results

--normal circumstances, possible worlds, time/place

That is, if a sequence contains the proposition that we took the train to Paris, we may omit propositions representing the facts that I perform actions to go to the station, that I buy a ticket, that I get into the train, that I read a book in the train, that I arrive in Paris, etc. All these facts are 'normal' constituents, i.e., occur in many or most possible worlds or courses of events where the 'main action' occurs. The macro-rule indeed defines the most 'important' proposition of a sequence, e.g., the proposition identifying the frame of which the other propositions are obligatory or possible parts. This importance shows in the *hierarchical structure* thus assigned by the macro-rules to the sequence, on the basis of the conceptual structure of the world (e.g., of actions)* and of our knowledge about it as represented in frames.

The fourth rule, CONSTRUCTION, is very similar, but in this case the proposition yielded by the macro-rules is not part of the text base itself. We now have a sequence of propositions which as a whole may be replaced by another proposition, indeed as in:

(11) I went to the station. I bought a ticket. I went to the train (. . .). The train arrived in Paris.

which may be replaced by the proposition 'I took the train to Paris', due to the fact that this proposition weakly or strongly implies the various propositions of the text base. As was spelled out above, the constituent propositions denote normal conditions, components, and consequences of the event described by the macro-proposition. In a sense, thus, the macro-proposition described the 'same' event, but only at a higher level of observation, comprehension, and function. When we read a book, we may have read a sequence like (11), but may simply have reduced this information by rule IV, and have stored that somebody took the train to Paris. By rule II, we may even generalize to the information that somebody 'went' to Paris.

The importance of the CONSTRUCTION rule, as well as of the GENERALIZATION rule, is that they define propositional information which is not as such expressed in the

----
*We have neglected here a systematic treatment of the structure of action and its relevance for action description, narrative, planning actions cognitively, and the comprehension of action discourse. Similarly, an action theory, from philosophy and philosophical logic, plays an important role in pragmatics. See for references, van Dijk (1975c, 1976b), especially to work done by von Wright, Davidson, Danto, Brennenstuhl, and others. For the operation of the macro-rules, it is necessary to know that the notions 'conditions', 'components' or 'consequences' of action mean exactly. Precise definitions cannot be given in this paper, but the intuitive idea about these notions will be sufficient to understand the conditions of the macro-rules.

discourse. Thus, it cannot be denied that 'I took the train to Paris' is information conveyed by (11), but this information is only present at a level of macro-interpretation. We see that macro-rules satisfy the more general principle of *entailment*: a macro-structure must be semantically or conventionally implied by the joint sequence of propositions of which it is a macro-structure. Secondly, the macro-rules may operate *recursively*: given a sequence of propositions on some macro-level, the rules may again apply and yield a macro-structure on a still higher level: my going to Paris may be a constituent of having a vacation in Europe.

There is however a constraint on recursion: in order to keep the macro-information as specific as possible, we always will stop at the lowest possible level of macro-structure, e.g., use the smallest superset or super-concept involved, and only when several propositions are the input to the rule. Thus, we do not generalize, by rule II, to 'I had a relation to Paris', or 'Somebody had a relation to some place'.

Another general constraint is that no proposition may be deleted which is a condition for the interpretation of a subsequent proposition, if this latter proposition itself is not deleted. Thus, macro-structures themselves must satisfy the normal conditions for linear connection and coherence.

Macro-structures may also be *expressed* by a discourse. Such a discourse will be called a *summary* of the original discourse. Depending on the level of the macro-structure expressed, *the* summary will be more or less general. When we speak of the summary of a discourse, we thereby mean the expression of *the* macro-structure of the discourse, namely, the top-most level of macro-structure.

Since a summary is a *natural discourse* and a macro-structure is a theoretical construct, the macro-structure is not fully expressed in surface structures of the summary; by the rules described above, some of the information may be kept implicit. Moreover, if a subject gives *his* summary of a discourse, there will be possible errors, biases, misinterpretation, e.g., due to pragmatic conditions of relevance assignment to certain propositions of the discourse.

Part of the macro-structure may also appear explicitly in the discourse itself, e.g., as in rule I and III, where macro-structural propositions are part of the text base. But, for reasons of cognitive processing, propositions which do not belong to the proper text base, e.g., those yielded by rules II and IV, also may be expressed in the discourse, e.g., as *announcements* and *partial summaries*. This fact provides one of the points where empirical evidence of macro-structures is to be found. Sentences expressing macro-structures in a discourse have very specific properties. For instance, they cannot normally be connected, e.g., in a complex sentence, with the other propositions of a sequence.

A clause or sentence expressing a macro-structural proposition, will be called a *topical* or *thematical* sentence. What is usually called a *topic-word* or *thematic word* or *key-word*, can now be defined as the lexical expression of one of the predicates of a macro-structural proposition, e.g., 'train, or 'travel', in the examples used above. As was pointed out earlier, such words only provide partial evidence of the global content of the discourse. Moreover, they *need* not be present at all, so that a mere key-word analysis may fail to be indicative of the global meaning of the discourse. Finally, we only know *that* a word has the function of a topic-word if we know what the semantic representation of the discourse is. Although, for instance, lexical frequency may indicate the relative importance of an underlying concept, this may only yield a *probabilistic hypothesis* about some part of the

Teun A. van Dijk

macro-structure of a discourse, not a rule-based description. First of all, lexical identity does not imply referential identity, which may determine discourse coherence. Second, macro-predicates may not be expressed directly by a topic-word. Third, even repeated lexical items with referential identity may merely have local importance. Fourth, lexical identity may indicate the conceptual range of a discourse (the 'topic' in the intuitive sense), but not as such define the specific information. Thus, in *all* discourses about oil, the word *oil* will probably be relatively frequent with respect to the overall frequencies of a corpus of discourses, but this does not yet indicate *what* the specific 'message' is 'about' the oil. For this we need propositional information, for which no overall frequencies can be available (the set of possible propositions of a language being infinite). *Titles* often express a part of the propositional macro-structure, and may therefore for practical reasons be used in information retrieval.

The earlier introduced notion *topic of discourse*, used to give conditions for linear connection and coherence, will now be identified with the notion of macro-structure as defined above. Thus, if we say that two propositions also are connected relative to a topic of discourse, we thereby mean that their connection not only is defined at the local, linear level, but also at the more global level, in the sense that they both should 'contribute' to the formation of a macro-structure. Hence, when we have the sequence:

(12) I went to the station and I bought a ticket

the two prepositions expressed also are connected by the fact that they denote two constituent facts which belong to the 'global fact' denoted by the macro-proposition 'I took the train'. Finally, a brief remark should be made about the macro-structural basis for discourse typologies.* Macro-structural propositions may also be assigned a (non-linguistic) function in a hierarchical structure defined by conventional rules, e.g., rules of narrative. Thus, a macro-proposition describing some initial state of an event or action, may be assigned the category of Setting. Similarly, for the categories of Complication, Resolution, Evaluation, and Moral, as defining simple narratives.** The point here is that isolated sentences or sequences, as such, cannot possibly have this function, because the constraints determining category assignment depend on the semantic properties of a macro-proposition. For example, the Complication category must represent an event but some sentences under this category in the story itself may well be state descriptions.

----

*The theory of discourse typologies (or 'genres') is little developed, although there is much traditional work on genres in literary scholarship. Sound linguistic approaches or advances in discourse studies in general are rare, especially because most attention is given to isolated discourse types, mostly narrative (in poetics and anthropology) to arguments (in philosophy) to everyday conversation and propaganda/advertisements (in the social sciences). See Gulich and Raible (1972) for a collection of papers. Our own paper on the subject, van Dijk (1972b), is inadequate.

**These categories come from Labov & Waletzky (1967 - and there are similar overall categories ('functions') of narrative being elaborated in the so-called 'structural analysis of narrative', begin in the late twenties in the Soviet Union, especially by Vladimir Propp, and rediscovered in the middle of the sixties in France (Bremond, Barthes, Greimas, Todorov, etc.). For references, see van Dijk (1972,1975c), and Culler (1975). See also the "On Narrative and Narratives" issue of *New Literary History* (Vol.VI, No. 2, Winter 1975), and Bremond (1973).

On the other hand, it may be the case that the macro-rules themselves operate under typological constraints: it is clear that an action in a narrative is more important globally speaking than a description of an object. Hence, the state descriptions will be deleted, and the action sequences mapped on action sequences in macro-structures. More in particular, we require the macro-rules to operate such that the specific type of the discourse is reflected in the macro-structure. This means that the summary of a story must itself also be a story. In retrieval, this allows the user of a system to know not only the 'content' of a discourse, but also the type of the original discourse.

In documentation this will prove to be particularly relevant in the storage of discourse with *argumentative* structures. Without the typological categories determining a macro-structure, we would perhaps have information concerning what the discourse is about, but not what (macro-) propositions count as premises, and which as conclusions. More generally, it will make an important difference in system use whether some information comes from a story told by somebody or from a scientific argument. It follows not only that a text grammar with its semantics is involved but also that there is a more general theory of discourse, in which the conventional categories of discourse types are defined, as well as their syntactical rules and semantic interpretation. No such explicit discourse typology exists at the moment, only the more descriptive results from such disciplines as rhetorics and poetics.

Macro-Structures and Information Science

In the preceding sections several conclusions from the theory of discourse structure have been drawn with respect to current or possible properties of procedures in information processing. Before we elaborate our suggestions, we may summarize the previous remarks as follows:

(i)   Information is semantic, and hence propositional; any content analysis, thus, must yield a sequence of propositions.

(ii)   The structure of propositions and of proposition sequences, in analysis, is determined by the syntactic surface structure of sentences; hence a syntax must be used, assigning structures which are semantically relevant.

(iii)   A sequence of propositions must satisfy the conditions of linear connection and coherence.

(iv)   Each discourse 'contains' information, i.e., propositions, which remain implicit, which are entailed by the explicit propositions (by meaning postulates and frame-knowledge), and which determine the linear coherence of the discourse.

(v)   Many elements of both surface structures and propositional structures of discourse are to be accounted for at pragmatic levels of description.

(vi)   Discourses with text structure also have global meanings, made explicit by macro-structures, which are proposition sequences obtained by macro-rules.

(vii)   Macro-information is entailed by the propositional text base of a discourse; it is necessary to assess the linear coherence of the discourse, to define the notion of a summary, a topical sentence and a topical word.

(viii)   Word-based and quantitative analysis of discourse surface structure are theoretically inadequate to assess semantic information, both at the linear, and especially at the global levels.

(ix)   The 'important' or 'relevant' information of a discourse is accounted for by the macro-structure of the discourse.

Since information science, e.g., in documentation, deals primarily with the processing of large numbers of complex discourses -- and not with the analysis of individual sentences, adequate models of processing must take into account the points summarized above. This means, first of all, that the discourses of documents are to be stored together with some form of semantic representation accounting for the 'content' of the discourses. Secondly, this content must be given not only in terms of the sequence of all propositions (implicit and explicit) of the discourse, but also in terms of various levels of macro-structure, which *organize* this sequence hierarchically, and which define various levels of *abstraction* or *reduction.*

Since it is assumed that the macro-structure defines what is semantically 'important' with respect to the discourse as a whole, and since it is practically impossible to store the full information of each document, the most adequate storage of complex semantic information is in the form of the most general macro-structure. In that case, by inverse application of the macro-rules at least part of the detailed information will be available by inference procedures.

Macro-structures are theoretical; in order to use them for practical purposes, they must be translated into some conventionally interpretable language, e.g., a specific version of a natural language. The result of such a translation is an (artificial) *summary* of the discourse.

Sets of summaries may in turn be subject to further organization, on the basis of their underlying propositions (which are macro-propositions of the corresponding discourses). We here arrive at what may be called the *pragmatic* aspects of information processing.* That is, not only is what is semantically 'important' with respect to the discourse to be accounted for, but at the same time the *functions* of the semantic information in communication between machine/system and man: questions asked, information given as answer, etc. In other words: one of the major co-determining constraints on information storage must be the possible use of the information in *most possible pragmatic contexts.* This implies that those macro-structures which most certainly will never be used will be deleted.

The pragmatic context involved in documentation processes is relatively simple: the machine/system for the most part will be required only to 'give information', i.e., to perform the speech act of *assertion.* In some cases, if further information is necessary in order to be able to provide the required information, the machine may ask further *questions.* The 'hearer' in the communicative context, i.e., the user of the documentation

----

*Although pragmatics is undoubtedly a major development in the humanities and the social sciences, it is neglected in this paper because, as is shown, the pragmatic properties of communication between machine/system and system user are rather simple. However, it should be stressed that the analysis of discourse itself must account for many types of phenomena which cannot be described in syntactic or pragmatic terms. For work in pragmatics, see Bar Hillel (1972), Wunderlich (1973), Cole and Morgan eds. (1975) and van Dijk, ed. (1975).

system, will for the most part perform the speech act of questioning. In order to be able to make *appropriate* assertions, the system must satisfy the conditions of appropriateness of assertions. First of all, it must only assert those propositions which the system-user does not yet know. The easiest way to do this is to take the user's question (and the propositions entailed by it) as the set of 'unknown and desired information' upon which the answer of the system must be based. Further the system will not give answers which are identical with general knowledge of the language and the world (i.e., lexicon and frames), because this knowledge is presupposed in all normal contexts of communication.

Now, the further organization of summaries of documents will depend on the *ranges* of the questions of system users, as defined by the conceptual structures of semantic space. Hence, in order to select a subset of summaries for further inspection of (propositional) information, we may use expressions for concepts from semantic space that are relevant for most system users. These expressions are given by some semi-artificial *descriptor language*. The concept expressed by a descriptor is, thus, a function characterizing a set of summaries, namely, the set of those summaries which use the concept in their macro-structure.

Note that besides a descriptor language for concepts, we may have a descriptor system for *names*. Since discourses not only have meanings but also refer to individuals, properties, and facts, in some possible world we also may want to retrieve information as a function of certain referents, e.g., well-known persons, places, or time periods. By using complex descriptor expressions, the set of relevant discourses may be further limited, but it should be emphasized again that descriptor sequences cannot possibly account for the (macro-) meaning of single discourses: they only define sets of summaries.

A more intelligent retrieval system based on these ideas need not have questions with underlying propositions directly answered by propositions of some stored summary. If we have the question "Where is Peter now?", and the macro-information from a story discourse about Peter's taking the train to Paris, a simple process of *inference* will try to derive an answer from the stored macro-information, with general meaning postulates and frames as additional premises. These processes are well-known from recent work in artificial intelligence and need not be discussed here.*

A system of the kind outlined roughly and informally above is not simple, of course. It must have a full morpho-syntax, if natural language questions may be asked. It must have a meaning and reference semantics, i.e., have a conceptual lexicon, and be able to construct artificial possible worlds in which individuals, properties, and facts can be located, such that these can be recovered for further reference. Furthermore, in order for the macro-rules to operate (as well as the rules of linear interpretation, for that matter), it must have a system of conventional knowledge about the actual world and worlds similar to it. For instance, it must know the abstract structure of events and actions in general, and of taking a bath, cashing a cheque, or eating in a restaurant in particular. If not, the system would be unable to recognize a sentence like "I took a beer with the pizza" as being part of the 'eating in an Italian restaurant' frame, and a sentence like "I finally paid at the cashier" as one of the final propositions in a store/shopping frame sequence. Without such a recognition, we would be unable to summarize a passage as "We ate in an Italian restaurant" or "I went shopping". Note, that not only are the frames themselves part of the system, but also the rules and constraints for handling them. Clearly, not all of the information which is, theoretically, part of a frame be used in the interpretation of a discourse, but only

----
*See, e.g., Minsky (1968), Schank and Colby (1973), Schank (1975).

that information which is a condition for the correct interpretation of following parts of the sequence. In order for somebody to understand my story about traveling to Paris, he need not actualize his full knowledge about 'how to take a cab', when it is said that I took a cab to the station. The only information relevant in that case is that a cab is a means of (public) transport.

On the basis of this frame system and the rules of frame application, and on the basis of the lexical meaning postulates, a set of inference rules operate in order to define derivational relations between (sequences of) propositions, as mentioned above.

Clearly, none of these components is ready yet at the moment, not even the syntax. This means that an adequate system of information retrieval can only be partial and theoretical at the moment, and handle fragmentary parts of discourse, world knowledge, and concepts. Yet, we at least know what such a system should contain in order to analyse and process discourse automatically, so that we know approximately which directions our work should take. Moreover, a certain number of basic concepts of information science can now be defined in a more or less explicit way, namely, in terms of macro-rules, macro-structures and their expressions in various languages or other systems.

Macro-Analysis: An Example

Finally, let us try to apply the theoretical framework sketched above in a more concrete analysis of a given discourse. Such an analysis, as our remarks in the previous section indicate, will of course be incomplete. We lack a morpho-syntactic analysis, and also a semantic analysis of the individual sentences/propositions of the discourse. We merely want to see whether and how the proposed macro-rules operate on an arbitrarily chosen discourse, and how these rules yield a macro-structure which can be expressed as an acceptable summary. Furthermore, this analysis will be informal, in the sense that no formal language is used to represent the semantic structure, to define the macro-rules, and to prove that the macro-structure is indeed entailed by the discourse.

The discourse we will give a partial macro-analysis of is a short paper from social psychology, "Bumper Stickers and the Cops", by F. K. Heussenstamm.* This discourse has been chosen first of all because it is short. Second, since information science deals predominantly with storage of scientific data, we have taken a scholarly paper. Third, the paper is relatively informal and not very complex. Fourth, papers in (social) psychology have a conventional typological structure. The paper has a clear political 'message', i.e., a moral conclusion drawn from a confirmed hypothesis. The original discourse is given in the *Appendix*.

The analysis will consist of various parts. First of all, we give a sequence of 'basic clauses', taken as approximate expressions of underlying propositions. From a logical point of view these propositions are still compound, but a full logical representation of the complete text would be too cumbersome for our purposes. In order to keep track of the introduced referents, we have added, between parentheses, letters representing the discourse referents involved, lower case letters for the local referents, upper case letters for the global referents. Similarly we have used the numbers of the propositions involved in those cases where a proposition is embedded in another proposition, mostly intensionally. Words which are underlined have a theoretical function (e.g., *of* denotes membership or subset); these may also occur in the discourse itself, e.g., as connectives. By underlining these connectives we denote connection with the previous proposition(s). We have not tried to translate the surface lexical items into basic conceptual predicates.

----
*From Swingle, P. G., ed., *Social Psychology of Everyday Life.* Penguin Books, 1973. Pp. 27-31.

In the second column of Table I, we indicate the *macro-rules* involved: I. DELETION, II. GENERALIZATION, III. SELECTION, and IV. CONSTRUCTION. M denotes topical propositions within the text itself, i.e., zero application of the rules. The resulting macro-proposition is mentioned at the end of each row of the column. The 'contribution' of a proposition to such a macro-proposition may be indirect. Thus if x is a condition for y, and y yields a macro-proposition Mz, then Mz is also specified after x.

The application of rule I does not yield a macro-proposition but a zero-unit, 0. Rule II, GENERALIZATION, often applies to several propositions at the same time. Such a sequence is indicated between angle brackets, where the numbers denote the corresponding propositions. If the generalizing proposition is part of the text itself, it has been underlined, e.g., II<55>.

Rules III and IV are accompanied by the specific type of condition involved, i.e., whether the substituted/deleted proposition denotes a CONDITION (COND), COMPONENT (COMP) or a CONSEQUENCE (CONS) of a fact denoted by a proposition in the text or a macro-proposition, respectively. In both cases the propositions relative to which the rules operate are also added between angled brackets.

A specific case involves those conditions, components or consequences having to do with mental acts and speech acts as they are related to denoted (macro-)acts. We use the abbreviation (MA) and (SA) to denote these specific conditions of deletion. Thus, if there is a fact, 'They did not like driving around with BPP bumper stickers', we may delete the fact that they 'said', 'reported', etc., so.

Sometimes several rules may apply on one proposition. In that case the alternatives are separated by a slash, e.g., II/III. In some cases we also have added the recursive application of rules, denoted by a vertical bar (|), mostly a combination of GENERALIZATION, preceding direct macro-expression: II|M, which means that a proposition is part of the macro-structure in its generalized form.

It will be obvious that our proposals for how the rules operate and the assumptions on the sufficiency of the rules are tentative and open to discussion. In some cases some information might have been taken up into the macro-structure, e.g., the fact that the immediate cause of the experiments were difficulties and complaints of students of the experimenter. We have however, integrated this information into macro-proposition M2, because the students were Black Panther Party members.

In Table II, we have listed the sequence of resulting macro-propositions. We see that the original number of propositions (164) is reduced to 29 macro-propositions, i.e., less than a fifth of the original number. This reduction is consistent with proposals for the macro-processing of a story of similar length given in an earlier paper. There it turned out that this number also has an empirical basis: it is what subjects recalling the discourse after some weeks still know.

The sequence of macro-propositions may be expressed by a natural language summary of the article, for which we have given a proposal in Table III. Such a summary has undergone the usual discourse output conditions of a grammar and pragmatics, including certain repetitions to make the discourse coherent, and the usual complex sentence constructions. Also certain lexical variants or combinations of propositions are possible here. A macro-structure, as we have argued above, is a function defining a set of *possible summaries*.

The summary is relatively long, but it contains all relevant information in the paper. Of course, there are 'degrees' of relevance, which means that macro-rules may again apply on this first-level macro-structure. These rules have been added in the second column of Table II, as described above. They yield a macro-structure underlying the summary given in Table IV, which conforms more to the *conventions* for giving a *short summary* of psychological papers. If we compare this summary with that of the author (or editor?) of the paper, we see that they run parallel for some time, but that the original summary accompanying the discourse is incomplete. It terminates with the major *finding* (result) of the experiment, and omits the conclusions.

Whereas the macro-structures organize and reduce the semantic structure of the discourse, there is another *superstructure* assigned to the discourse. It is the conventional structure of a paper in (experimental) social psychology, involving such categories as 'Introduction', 'Problem', 'Hypothesis', 'Experiment', 'Discussion', 'Conclusion', where the 'Experiment' category may again consist of 'Design', 'Subjects', 'Method', etc. A tentative proposal for this kind of hierarchical conventional super-structure for this text is given in Table V, together with the sequences of macro-propositions organized by these categories. It is important to note that the relative relevance of information in the paper is determined by the conventional structure of reports of psychological experiments.

Concluding Remarks

In this paper we have tried to show that information science in general and the theory of documentation in particular may profit from some recent developments in linguistics, psychology, and artificial intelligence with respect to complex semantic information processing, especially of discourse. It has been shown that the notion of 'information in a document' should be constructed in terms of an explicit semantics, both intensional and referential, and that important notions such as 'content', 'topic' or 'theme', 'abstract' or 'summary', and 'key-word' can be appropriately defined in that framework.

More in particular it has been argued that information storage and retrieval must be based on operations of organization and reduction of complex semantic information. These operations are modelled by a theory of macro-structures, which represent the global semantic structure of discourses, and which are obtained from the propositional sequence of a discourse by a set of macro-rules. Storage and retrieval, both natural and artificial, is a function of these macro-structures as assigned to the discourse during analysis/input. Some conditions were formulated for the linear connection and coherence of complex sentences and discourse in general, primarily based on both intensional and referential relations between facts in possible worlds. The notion of 'topic of discourse', which was assumed to co-determine linear coherence, was found to be identical with (partial) macro-structure, which links the linear coherence with the global, overall coherence of discourse. Besides these various levels of semantic micro- and macro-structures of discourse, there are also global conventional rules determining the hierarchical organization of different types of discourses, e.g., of narratives.

In this paper we have, in particular, tried to apply this theoretical sketch in the concrete analysis of a scholarly paper from social psychology, by establishing a tentative sequence of propositions for the discourse, letting macro-rules operate on (sub-)sequences, thus yielding a macro-structure. This macro-structure can be expressed in natural language discourse as a summary of the paper. Further application of the macro-rules may lead to shorter and more general summaries.

Perspective Paper: Complex Semantic Information Processing

This 'summary' of our own paper does not mention the numerous problems remaining. To be sure, both our rules and the constraints determining their application are very tentative. Strictly speaking, we are still in the 'intuitive' stage of theory formation, because we do not have sufficient sub-systems to have the rules work algorithmically, which would be a condition for computer implementation. Although we have tried to show that a certain number of approaches (quantitative, lexical/morpho-syntactic) must be theoretically inadequate because they do not reconstruct the notion of 'information', we cannot possibly offer at the moment a directly applicable alternative. Language structure and language use simply are not so simple. This truism just means that the theoretical framework to be constructed for information processing purposes will have a large number of highly complex sub-theories. Besides the mentioned morpho-syntactic and semantic components, of which only small fragments are available at the moment, we first of all need an appropriate lexicon, codifying the conventional conceptual meaning structures as they are organized lexically by a given language.

Secondly, parallel to such a lexicon, we need (sub-)systems representing our conventional knowledge of the world, and insight into the acquisition and transformations of these so-called *frames*, and in particular into the ways such frames operate in complex behavior in general, and in discourse comprehension in particular. Only the first steps have been made in recent artificial intelligence in the elaboration of a frame theory. It is obvious that information science, which is essentially concerned with storage/retrieval, i.e., organizational, problems with respect to our ('collective') knowledge, could immediately profit from advances in that area of research, and by its own investigations and experience contribute to it. In particular, the role of world knowledge in the interpretation and storage of natural language discourse may thus become a challenging problem both to information science and to linguistics, psychology, artificial intelligence, and computer science. This dialogue between what now are being called various 'cognitive sciences' may provide the basis for a solution of the still more complex but crucially relevant problems of social information processing as it determines social interaction.

References

Ballmer, T. T. "Macrostructures". In van Dijk, ed., *Pragmatics of Language and Literature.* Amsterdam, North Holland, 1975, Pp. 1-22.

Bar-Hillel, Y., ed. *Pragmatics of Natural Language.* Dordrecht, Reidel, 1972.

Bauman, R., and Scherzer, J., eds. *Explorations in the Ethnography of Speaking.* London, Cambridge University Press, 1974.

Bobrow, D. G., and Collins, A., eds. *Representation and Understanding: Studies in Cognitive Science.* New York, Academic Press, 1975.

Bremond, C., ed. *Sémiotique Narrative et Textuelle.* Paris, Larousse, 1973.

Charniak, E. Towards a Model of Children's Story Comprehension. AI TR-266, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, 1972.

Cicourel, A. V. *Cognitive Sociology.* Harmondsworth, Penguin Books, 1973. New York, Academic Press, 1975.

Teun A. van Dijk

Cresswell, M. J. *Logics and Languages.* London, Methuen, 1973.

Culler, J. *Structuralist Poetics.* London, Routledge and Kegan Paul, 1975.

van Dijk, T. A. *Some Aspects of Text Grammars.* The Hague, Mouton, 1972. (a)

van Dijk, T. A. "Foundations for Typologies of Texts." *Semiotica,* 1972, 6, 297-323. (b)

van Dijk, T. A. "Relevance in Logic and Grammar." *Proceedings, International Congress of Relevance Logics,* St. Louis, 1974.

van Dijk, T. A. "Recalling and Summarizing Complex Discourse." University of Amsterdam, 1975. (a)

van Dijk, T. A. "Narrative Macro-structures." (Paper contributed to the Linguistics and Poetics Colloquium of the University of Essex, 1975). To appear in *PTL.* (b)

van Dijk, T. A. "Action, Action Description, Narrative." *New Literary History,* 1975, VI, 273-294. (c)

van Dijk, T. A. "Connectives in Text Grammar and Text Logic." (Paper for the International Text Linguistics Colloquium, Kiel, 1973). In van Dijk, T.A., and Petöfi, J.S., eds., *Grammars and Descriptions.* New York, Berlin, De Gruyter, 1976. (a)

van Dijk, T. A. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse.* University of Amsterdam, 1976. (b) (Probably to be published by Longmans, London.)

van Dijk, T. A. "Macro-structures and Cognition." (Paper contributed to the Carnegie-Mellon Symposium on Cognition, May 1976.) University of Amsterdam, 1976.(c)

van Dijk, T. A., ed. *Pragmatics of Language and Literature.* Amsterdam, North Holland, 1975.

van Dijk, T. A., and Petofi, J. S., eds. *Grammars and Descriptions.* New York, Berlin, De Gruyter, 1976.

Dressler, W. U., ed. *Trends in Text Linguistics.* New York, Berlin, De Gruyter, 1977.

Dressler, W. U., and Schmidt, S. J. *Textlinguistik: Kommentierte Bibliographie.* Munich, Fink, 1973.

Freedle, R. O., and Carroll, J. B., eds. *Language Comprehension and the Acquisition of Knowledge.* Washington, D.C., Winston, 1972.

Gerbner, G., et al., eds. *The Analysis of Communication Content.* New York, Wiley, 1969.

Gulich, E., and Raible, W., eds. *Textsorten.* Frankfurt, Athenaeum, 1972.

Perspective Paper: Complex Semantic Information Processing

Hintikka, K. J. J., Moravcsik, J. M. E., and Suppes, P., eds. *Approaches to Natural Language.* Dordrecht, Reidel, 1973.

Holsti, O. R. *Content Analysis for the Social Sciences and the Humanities.* Reading, Mass., Addison Wesley, 1969.

Keenan, E. L., ed. *Formal Semantics of Natural Language.* London, Cambridge University Press, 1975.

Kintsch, W. The *Representation of Meaning in Memory.* Hillsdale, N. J., Erlbaum, 1974. (Distributed by Wiley, New York.)

Kintsch, W., and van Dijk, T. A. "Recalling and Summarizing Stories." University of Colorado, 1975. [French version: "Comment on se Rappelle et on Resume des Histoires." *Langages,* 1975, 40, 98-116.]

Kintsch, W., and van Dijk, T. A. "Cognitive Psychology and Discourse." University of Colorado, 1976. (Elaborated version of Kintsch and van Dijk, 1975.) [To be published in Dressler, W.U., ed., *Trends in Text Linguistics.* New York, Berlin, De Gruyter, 1977.

Labov, W., and Waletzky, Y. "Narrative Analysis: Oral Versions of Personal Experience." In Helm, J., ed., *Studies in the Verbal and Visual Arts.* Seattle, Washington University Press, 1967. Pp. 12-44.

Meyer, B. F. *The Organization of Prose and its Effects on Memory.* Amsterdam, North Holland, 1975.

Minsky, M. "A Framework for Representing Knowledge." In Winston, P., ed. *The Psychology of Computer Vision.* New York, McGraw Hill, 1975.

Minsky, M., ed. *Semantic Information Processing.* Cambridge, Massachusetts, MIT Press, 1968.

Montague, R. *Formal Philosophy: Selected Papers of Richard Montague.* Edited by R.H. Thomason. New Haven, Yale University Press, 1974.

Norman, D. A., and Rumelhart, D. E., eds. *Explorations in Cognition.* San Francisco, Freeman, 1975.

Petöfi, J. S., and Rieser, H., eds. *Studies in Text Grammar.* Dordrecht, Reidel, 1973.

Schank, R. C., ed. *Conceptual Information Processing.* Amsterdam, North Holland, 1975.

Schank, R. C., and Colby, K. M., eds. *Computer Models of Thought and Language.* San Francisco, Freeman, 1973.

Sparck Jones, K., and Kay, M. *Linguistics and Information Science.* New York, Academic Press, 1973.

Sudnow, D., ed. *Studies in Social Interaction.* New York, Free Press, 1972.

Thorndyke, P. W. "Cognitive Structures in Human Story Comprehension and Memory." Ph.D. Thesis, Stanford University, 1975.

Turner, R.H., ed. *Ethnomethodology.* Harmondsworth, Penguin Books, 1973.

Wunderlich, D., ed. *Linguistische Pragmatik.* Frankfurt, Athenaeum, 1972.

APPENDIX

Bumper Stickers and the Cops

(From: P.G. Swindle, ed. *Social Psychology of Everyday Life.* Harmondsworth, Penguin Books, 1973. Pp. 27-31)

3. Bumper stickers and the cops *F. K. Heussenstamm*

> During a discussion about a student group of Black Panther Party members who had received so many traffic citations that they were in danger of losing their licenses, it was discovered that all had Panther Party stickers glued to their bumpers. The Panthers' claims of police harassment were put to the test by having 15 drivers with no traffic violations for the preceding 12 months attach Black Panther stickers to the rear bumpers of their cars. With the bumper stickers on their cars, the students received a total of 33 citations in 17 days.

A series of violent, bloody encounters between police and Black Panther Party members punctuated the early summer days of 1969. Soon after, a group of black students I teach at California State College, Los Angeles, who were members of the Panther Party, began to complain of continuous harassment by law enforcement officers. Among their many grievances, they complained about receiving so many traffic citations that some were in danger of losing their driving privileges. During one lengthy discussion, we realized that all of them drove automobiles with Panther Party signs glued to their bumpers. This is a report of a study that I undertook to assess the seriousness of their charges and to determine whether we were hearing the voice of paranoia or reality.

Recruitment advertising for subjects to participate in the research elicited 45 possible subjects from the student body. Careful screening thinned the ranks to 15 -- 5 black, 5 white, and 5 of Mexican descent. Each group included three males and two females. Although the college enrolls more than 20,000 students (largest minority group numbers on the west coast) it provides no residential facilities; all participants, of necessity, then, traveled to campus daily on freeways or surface streets. The average round trip was roughly 10 miles, but some drove as far as 18 miles. Eleven of the 15 had part-time jobs which involved driving to and from work after class as well.

All participants in the study had exemplary driving records, attested to by a sworn statement that each driver had received no "moving" traffic violations in the preceding 12 months. In addition, each promised to continue to drive in accordance with all in-force Department of Motor Vehicles regulations. Each student signed another statement to the effect that he would do nothing to "attract the attention" of either police, sheriff's deputies, or highway patrolmen -- all of whom survey traffic in Los Angeles county. The participants declared that their cars, which ranged from a "flower child" hippie van to standard American makes of all types, had no defective equipment. Lights, horns, brakes,

and tires were duly inspected and pronounced satisfactory.

The appearance of the drivers was varied. There were three blacks with processed hair and two with exaggerated naturals, two white-shirt-and-necktie, straight Caucasians and a shoulder-length-maned hippie, and two mustache- and sideburn-sporting Mexican-Americans. All wore typical campus dress, with the exception of the resident hippie and the militant blacks, who sometimes wore dashikis.

A fund of $500 was obtained from a private source to pay fines for any citations received by the driving pool, and students were briefed on the purposes of the study. After a review of lawful operation of motor vehicles, all agreed on the seriousness of receiving excessive moving traffic violations.   In California, four citations within a 12-month period precipitates automatic examination of driving records, with a year of probation likely, or, depending on the seriousness of the offenses, suspension of the driver's license for varying lengths of time.   Probation or suspension is usually accompanied by commensurate increases in insurance premiums. Thus, the students knew they were accepting considerable personal jeopardy as a condition of involvement in the study.

Bumper stickers in lurid day-glo orange and black, depicting a menacing panther with large BLACK PANTHER lettering, were attached to the rear bumper of each subject car, and the study began. The first student received a ticket for making an "incorrect lane change" on the freeway less than two hours after heading home in the rush hour traffic. Five more tickets were received by others on the second day for "following too closely", "failing to yield the right of way", "driving too slowly in the high-speed lane of the freeway", "failure to make a proper signal before turning right at an intersection", and "failure to observe proper safety of pedestrians using a crosswalk". On day three, students were cited for "excessive speed", "making unsafe lane changes" and "driving erratically". And so it went every day.

One student was forced to drop out of the study by day four, because he had already received three citations. Three others reached what we had agreed was the maximum limit -- three citations -- within the first week. Altogether, the participants received 33 citations in 17 days, and the violations fund was exhausted.

Drivers reported that their encounters with the intercepting officers ranged from affable and "standard polite" to surly, accompanied by search of the vehicle. Five cars were thoroughly gone over and their drivers were shaken down. One white girl, a striking blonde and a member of a leading campus sorority, was questioned at length about her reasons for supporting the "criminal activity" of the Black Panther Party. This was the only time that an actual reference to the bumper stickers was made during any of the ticketings. Students, by prior agreement, made no effort to dissuade officers from giving citations, once the vehicle had been halted.

Pledges to Drive Safely

Students received citations equally, regardless of race or sex or ethnicity or personal appearance. Being in jeopardy made them "nervous" and "edgy", and they reported being very uncomfortable whenever they were in their automobiles. After the first few days, black students stopped saying "I told you so", and showed a sober, demoralized air of futility. Continuous pledges to safe driving were made daily, and all expressed increasing incredulity as the totals mounted.   They paid their fines in person immediately after receiving a citation. One student received his second ticket on the way to pay his fine for

the first one.

No student requested a court appearance to protest a citation, regardless of the circumstances surrounding a ticketing incident. When the investigator announced the end of the study on the 18th day, the remaining drivers expressed relief and went straight to their cars to remove the stickers.

Some citations were undoubtedly deserved. How many, we cannot be sure. A tightly designed replication of this study would involve control of make and year of cars through the use of standard rented vehicles of low-intensity color. A driving pool of individuals who represented an equal number of both extreme-left and straight-looking appearance with matched age-range could be developed. Drivers could be assigned at random to preselected, alternate routes of a set length. Both left-wing and right-wing bumper sticker could also be attached at random after drivers were seated in their assigned vehicles and the doors sealed. In this way, no subject would know in advance whether he was driving around with "Black Panther Party" or "America Love It Or Leave It" on his auto. This would permit us to check actual driving behavior in a more reliable way. We might also wish to include a tape recorder in each car to preserve the dialogue at citation incidents.

No More Stickers

It is possible, of course, that the subject's bias influenced his driving, making it less circumspect than usual. But it is statistically unlikely that this number of previously "safe" drivers could amass such a collection of tickets without assuming real bias by police against drivers with Black Panther bumper sticker.

The reactions of the traffic officers might have been influenced, and we hypothesize that they were, by the recent deaths of police in collision with Black Panther Party members. But whatever the provocation, unwarranted traffic citations are a clear violation of the civil rights of citizens and cannot be tolerated. Unattended, the legitimate grievances of the black community against individuals who represent agencies of the dominant society contribute to the climate of hostility between the races at all levels and predispose victims to acts of violent retaliation.

As a footnote to this study, I should mention that Black Panther bumper stickers are not seen in Los Angeles these days, although the party has considerable local strength. Apparently members discovered for themselves the danger of blatantly announcing their politics on their bumpers and have long since removed the "incriminating" evidence.

Table I. Sequence of Propositions

1. There was a series of encounters(a) between police(A) and members of the Black Panther Party(B)	M	M1
2. (a) were violent	M	M1
3. (a) were bloody	II<u>2</u>>	M1
4. (1) was in 1969	I	0
5. (1) was in early summer	I	0
6. Soon after (1) a group of students(C) began to complain of (11)	III<11>CONS	M2
7. (C) were black	M	M2
8. I(D) teach(C) at California State College(b)	I	0
9. (b) is in Los Angeles	I	0
10. (C) were members of (B)	M	M2
11. Law enforcement officers(<u>of</u> A) harassed (C)	M	M2
12. (11) was continuous	I	0
13. (C) had many grievances(c) I 0		
14. One of (c) was a complaint by (C) about (15)	III<15> CONS(SA)	M3
15. (C) received many traffic citations(E)	M	M3
16. (15) is a danger of losing (C)'s driving privileges	III<15> CONS	M3
17. WE (C <u>and</u> D) realized that (18-19)	III<18,19>CONS(MA)	M4
18. All of (C) drove automobiles(d)	M	M4
19. (d) had Panther Party signs(F) glued to their bumpers	M	M4
20. <u>This</u> is a report of a study(G)	III<21> CONS(SA)	M5
21. I(D) undertook (G) for (22) and (23)	III<22,23> COND	M5
22. I(D) assess seriousness of (C)'s charges	II<23>	M5
23. I(D) determine whether we (C <u>and</u> D) were hearing the voice of paranoia or reality	II<23>	M5
24. Recruitment(e) elicited 45 possible subjects from the student body(f)	IV<M6> COMP	M6
25. (e) was advertising for subjects to participate in the research	III<24> COND	M6
26. Screening(f) thinned the ranks to 15(H)	IV<M6>	M6
27. 5 of (H) were black(g)	II<27-31>	M11

28. 5 of (H) were white(h)	II⟨27-31⟩	M11
29. 5 of (H) were of Mexican descent(i)	II⟨27-31⟩	M11
30. Each (g)(h)(i) included three males	II⟨27-31⟩	M11
31. Each (g)(h)(i) included three females	II⟨27-31⟩	M11
32. The college(j) enrolls more than 20,000 students	I	0
33. (j) has the largest minority group numbers on the West Coast	I	0
34. (j) yet provides no residential facilities	III⟨35⟩ COND	M7
35. All (H) necessarily traveled to campus	II	M7
36. (35) was daily	I	0
37. (35) was on freeways or surface streets	III⟨35⟩ COMP	M7
38. The average round trip was roughly 10 miles	I	0
39. But some of (H) drove as far as 18 miles	I	0
40. 11 of the 15(H) has part-time jobs(1)	I	0
41. (1) involved driving to and from work after class as well	I	0
42. All participants(H) in the study had exemplary driving records(m)	M	M8
43. (m) were attested by a sworn statement that (44)	III⟨42⟩ CONS(SA)	M8
44. Each driver (of (H)) had received no "moving" traffic violations in the preceding 12 months	M	M9
45. In addition each (H) promised that (46)	III⟨46⟩ COND	M10
46. (H) continues to drive in accordance with all in-force Department of Motor Vehicles regulations	M	M10
47. Each student (H) signed another statement to the effect that (48)	III⟨48⟩ COND(SA)	M10
48. Each (H) will do nothing to "attract the attention" of either police, sheriff's deputies, or highway patrolmen (I)	II|M	M10
49. All of (I) survey traffic in Los Angeles county	I	0
50. The participants (H) declared that (51)	III⟨51⟩ COND(SA)	M13
51. Their(H) cars(J) have no defective equipment	II|M	M13

152

52. (J) range from a "flower child" hippie van to standard American makes of all types	I/II	0
53. Lights, horns, brakes, and tires(n) were duly inspected	III<54> COND	M13
54. (n) were pronounced satisfactory	M	M13
55. The appearance of the drivers(H) was varied	M	M11
56. Three blacks(g) had processed hair	II<55>	M11
57. Two(of (H)) had exaggerated naturals	II<55>	M11
58. Two(of (H)) had white-shirt-and-necktie (o)	II<55>	M11
59. (o) were straight Caucasians	II<55>	M11
60. One(of (H)) was a shoulder-length-maned hippie	II<55>	M11
61. Two(of (H)) were mustache- and sideburn-sporting Mexican-Americans	II<55>	M11
62. All(H) wore typical campus dress	II<55>	M11
63. Except one (of (H)) and the militant blacks (p)	II<55>	M11
64. (p) sometimes wore dashikis	II<55>	M11
65. A fund of $500 was obtained from a private source for (66)	III<66> COND/ I	M19 0
66. (H) pays fines (K) for any citations(L) received by the driving pool(H)	IV< M19> COND	M19
67. Students(H) were briefed on the purposes of the study(G)	IV<M5> COND	M5
68. The lawful operation of motor vehicles was reviewed	III<48> COND	M10
69. (Then) all (of (H)) agreed on the seriousness of (70)	M	M12
70. (H) receive excessive moving traffic violations	II<69-78>|M	M12
71. In California 4 citations(of L) within a 12-month period precipitates automatic (72)	III<73> COND	M12
72. Driving records are examined	II<69-78>	M12
73. ((71) cause) likely year of probation(q)	II<69-78>	M12
74. Or supension of the driver's license (r) for varying lengths of time	II<69-78>	M12
75. (q) depends on the seriousness of the offenses	III<73> COND	M12

153

76. Probation(q) or suspension(r) is usually accompanied by commensurate increases in insurance premiums	II<69-78>	M12
77. Thus, the students(H) knew (that (78))	III<78> COND	M12
78. (H) were accepting considerable personal jeopardy	II<69>	M12
79. (78) was a condition of involvement in the study(G)	III<M6> COMP	M6
80. Bumper stickers(M) were attached to the rear bumper(N) of each subject car(J)	IV<M14> COMP	M14
81. (M) were in lurid day-glo orange and black	I	O
82. (M) were depicting a menacing panther with large BLACK PANTHER lettering	II<80-82>	M14
83. And the study(G) began	IV<M14-21> COMP	M5
84. The first student(s) (of H) received a ticket(L') for making an "incorrect lane change" on the freeway	IV<M15> COMP	M20
85. (84) was less than 2 hours after (86)	I	O
86. (s) headed home in the rush hour traffic	I	O
87. 5 more (of L') were received by others (of H) on the second day for "following too closely", "failing to yield the right of way", "driving too slowly in the high speed lane of the freeway", "failure to make a proper signal before turning right at an intersection", and "failure to observe proper safety of pedestrians using a crosswalk"	IV<M15> COMP	M15
88. On day three, students (of (H)) were cited for "excessive speed", "making unsafe lane changes", and "driving erratically"	IV<M15> COMP	M15
89. And so (87, 88) it went every day	II	M15
90. One student(t) (of (H)) was forced to (91)	II<90-94>	M16
91. (t) drop our of the study(G) by day four	II<90-94>	M16
92. (t) had already received 3 citations within the first week caused (90, 91)	III<90-91> COND	M16
93. 3 others (of (H)) reached the maximum limit(u)	II<90-91>	M16
94. We((D) and (H)) had agreed that (u) was 3 citations	IV<M16> COND(SA)	M16
95. Altogether, the participants(H) received 33 citations(L) in 17 days	M	M17

154

96. And (so) the violation fund was exhausted	IV<M17> CONS	M17
97. Driver(H) reported that (98)	III<98> CONS(SA)	M20
98. Their(H) encounters with the intercepting officers ranged from affable and "standard polite" to surly(v)	II<98-106>	M20
99. (v) was accompanied by search of the vehicle	II<98-106>	M20
100. 5 cars (of (J)) were thoroughly gone over	II<98-106>	M20
101. Their(J) drivers(of H) were shaken down	II<98-106>	M20
102. 1 white girl(w) (of H) was questioned at length about her reasons for (103)	II<98-106>	M20
103. (w) support the "criminal activity" of the Black Panther Party(B)	II<98-106>	M20
104. (w) was a striking blonde I O		
105. (w) was a member of a leading campus sorority	I	O
106. This (102) was the only time that an actual reference to the bumper stickers (M) was made during any of the ticketings	III<102> COND	M20
107. Students(H) by prior agreement made no effort to dissuade officers(x) (of I) from (108)	IV<M19> COMP	M19
108. (x) gave citations (of L)	M	M15
109. (After) the vehicle (of J) had been halted (106)	III<108> COND	M15
110. Students (of H) received citations (of L) equally	II /M	M18
111. (110) was regardless of race or sex or ethnicity or personal appearance	II	M18
112. Being in jeopardy(78) made them(H) "nervous" and "edgy"	I/II	M21
113. They(H) reported (114)	III<114> CONS(SA)	M21
114. (H) were very uncomfortable whenever they were in their automobiles (J)	II/IV<M14> CONS	M21
115. After the first few days black students(C) stopped saying "I told you(D) so"	I	O
116. (C) showed a sober, demoralized air of futility	I	O
117. Continuous pledges to drive safely were made daily	IV<M10> COMP(SA)	M10
118. All (of H) expressed increasing incredulity	IV<M17> CONS(SA)	M17

155

119. The totals mounted	II	M17
120. They(H) paid their(H) fines	M	M19
121. (120) was immediately after receiving a citation (of L)	III<120> CONS/ III<110> COMP	M19 M10
122. One student(y) (of H) received his(y) second ticket (of L')	IV< M15> COMP	M15
123. (122) was on the way to pay his(y) first one (of L')	I	0
124. No student (of H) requested a court appearance to (125)	III<125> COND(SA)	M19
125. (of H) protest a citation (of L)	M/IV< M10> COMP	M19
126. (124, 125) was regardless of the circum-stances surrounding a ticketing incident	I	0
127. When the investigator (D) announced the end of the study (G) on the 18th day (128)	IV<M5> COMP(SA)	M5/ M21
128. The remaining drivers(z) (of H) expressed relief	III<114> CONS(SA) IV< M21> CONS	M21
129. (z) and went straight to their cars to (130)	IV< M21> CONS	M21
130. (z) removed the stickers	IV< M21> CONS	M21
131. Some citations (of L) were undoubtedly deserved	M	M22
132. We (D) cannot be sure how many (of L)	I	0
133. A replication of this study (G) would involve (134-147)	M	M23
134. Make and year of cars are controlled through 135	II/IV<M23> COMP	M23
135. Use is made of standard rented vehicles of low-intensity color	II/IV<M23> COMP	M23
136. A driving pool of individuals(aa) could be developed	II/IV<M23> COMP	M23
137. (aa) represent equal number of both extreme-left and straight-looking appearance with matched age-range	II/IV<M23> COMP	M23
138. Drivers(aa) could be assigned at random to preselected alternate routes of a set length	II/IV<M23> COMP	M23
139. Both left-wing and right-wing bumper stickers could also be attached at random after (140-141)	II/IV<M23> COMP /III<139-141> COND	M23

140. Drivers(aa) are seated in their assigned vehicles	II/IV<M23> COMP	M23
141. And the doors are sealed	II/IV<M23> COMP	M23
142. In this(139) way no subject (of aa) would know in advance whether (143) or (144)	II/IV<M23> COMP	M23
143. (of aa) is driving around with "Black Panther Party" on his auto	II/IV<M23> COMP III<145> COND	M23
144. (of aa) is driving around with "America Love It or Leave It" on his auto	II/IV<M23> COMP III<145> COND	M23
145. This (142) would permit us (D) to check actual driving behavior in a more reliable way	II/IV<M23> COMP	M23
146. We(D) might also wish to include a tape recorder in each car to (147)	II/IV<M23> COMP III<147> COND	M23
147. We(D)) preserve the dialogue at citation incidents	II/IV<M23> COMP	M23
148. It is possible, of course, that the subject's (of H) bias(ab) influenced his driving(ac)	II<150>	M24
149. (ab) make it(ac) less circumspect than usual	III<148> CONS	M24
150. But it is statistically unlikely that the number of previously "safe" drivers (G) could amass such a collection of tickets (L') without 151	II|M	M24
151. (D) assumes real bias by police against drivers with Black Panther bumper stickers	M	M25
152. The reactions of the traffic officers (of I) might have been influenced	M	M26
153. And we(D) hypothesize that (154)	III<154> CONS(SA)	M26
154. (of I) were(influenced) by the recent deaths of police (of A) in collision with Black Panther Party members(B) (1)	M	M26
155. But whatever the provocation, unwarranted traffic violations (of L) are a clear violation of the civil rights of citizens	II/M	M27
156. (155) cannot be tolerated	III<155> CONS	M27
157. If unattended(ad) then (158)	II	M28
158. The legitimate grievances (ad) of the black community against individuals who represent agencies of the dominating society contribute to the climate of hostility between the races at all levels and (159)	II	M28

157

159. (ad) predispose victims to acts of violent retaliation	II	M28
160. As a footnote to this study(G) I(D) should mention (161)	III<161> CONS(SA)	M29
161. Black Panther bumper stickers (of F) are not seen in Los Angeles these days	M	M29
162. But the party has considerable local strength	I	0
163. Apparently members (B) discovered for themselves(B) the danger of blatantly announcing their politics on their bumpers	III<161> COND III< M15> CONS	M29
164. (of B) have long since removed the "incriminating" evidence	III<161> COND	M29

Table II. Sequence of Macro-Propositions

1. There were violent encounters between police(A) and Black Panther Party members(B)	M
2. Soon after (1) (A) harassed Black Panther Party members(C) by (3)	I/III<1> CONS
3. (A) gave many traffic citations(E) to (C)	M
4. Probably (3) resulted from Black Panther Party bumper stickers on the cars of (C)	M
5. We made a study with the purpose to confirm (4)	M
6. A group of 15 students(G) was selected	M
7. (G) normally traveled by car	III<M8> COND
8. (G) had exemplary driving records	M
9. (G) had not received citations for 12 months	III<M8> COND
10. (G) promised to drive according to the rules and not to attract police attention	M
11. (G) had various appearances and backgrounds	M
12. (G) realized the seriousness of the experiment	I/III<M15> CONS
13. The cars (J) were found OK	III<M10> COND
14. (G) drove around with BPP bumper stickers	M
15. (G) were regularly fined for various minor offenses	M
16. Some of (G) had to drop out	I
17. (G) received 33 citations in 17 days	M
18. All (of G) were fined uniformly	M
19. All fines were paid without protest	M
20. The police behaved differently	I
21. (G) did not like (14)	I/III<M15> CONS
22. Most citations were not deserved	M
23. A replication with better control of appearance of subjects, cars, and of behavior of subjects would be necessary	II
24. But it is unlikely that good drivers suddenly become worse drivers	M
25. Hence police bias against drivers with Black Panther Party bumper stickers	M

26. (25) is probably caused by (1)	M
27. Yet (25) is an unacceptable violation of civil rights	II
28. (27) may cause violent behavior of (B)	III<27> CONS
29. (25) caused that Black Panther Party bumper stickers no longer are on cars of (B)	M

Table III. A Possible (Long) Summary

After violent encounters with members of the Black Panther Party, the police harassed them by giving them many traffic citations, probably due to Black Panther Party bumper stickers on their cars.

An experiment was carried out in order to see whether this assumption of biases in traffic citations was correct.

A group of 15 students was selected who usually traveled by car, and who had excellent driving records, not having had traffic citations for at least a year.

They promised to drive according to the rules and not to attract the attention of the police.

Their appearance and background were varied.

They realized that their task was serious.

After inspection of their cars, they drove around in their cars with Black Panther Party bumper stickers.

As expected, they were indeed fined for various minor reasons.

Some of the subjects had to drop out.

Ss received a total of 33 citations in 17 days without marked differences for individual subjects.

The behavior of the police was different.

The fines were paid without protest.

The subjects did not like their task much.

It should be concluded that most citations were not deserved.

Although a replication with better control of cars, subjects, and driving behavior would perhaps be necessary, it is unlikely that good drivers suddenly become bad drivers.

It follows that the police indeed has a bias in giving citations to Black Panther Party members.

Although this may be understandable by the recent riots, such a bias is a violation of civil rights, and may cause further violent behavior of the Black Panthers.

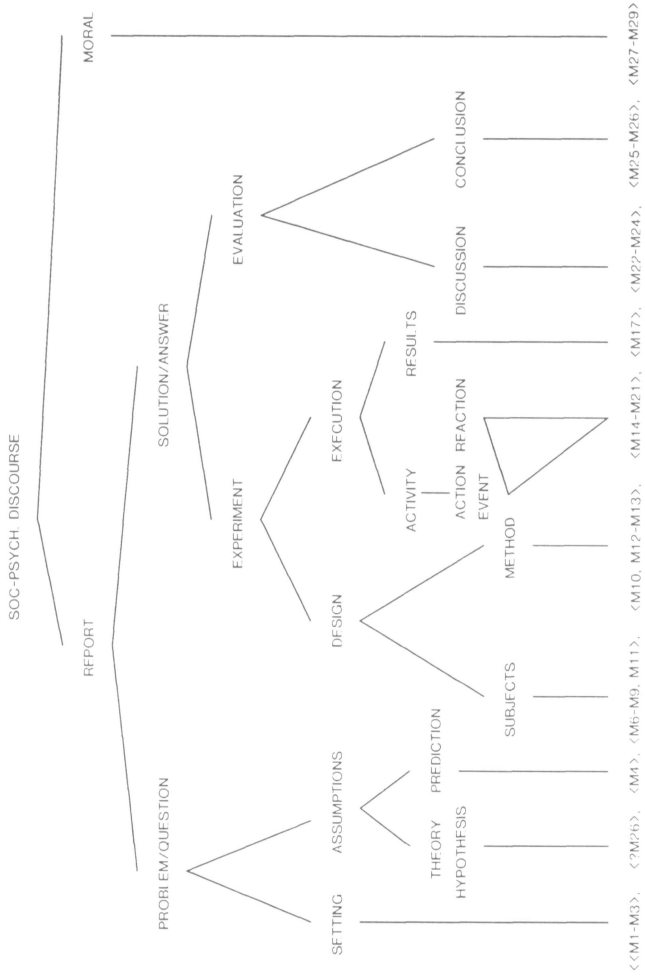Since these had noticed the effect of the bumper stickers, they no longer drive with them.

Table IV. A Possible (Short) Summary

An experiment was carried out in order to test the hypothesis that membership in the Black Panther Party, as shown by bumper stickers, leads to biases in giving traffic citations by the police, possibly as a result of recent riots between the Black Panthers and the police.

Subjects with varying appearances and backgrounds were selected who all had very good driving records. After instruction not to attract the attention of the police and to drive according to the rules, they drove around with Black Panther Party bumper stickers. As expected they (N=15) received many citations (33 in 17 days), independent of individuals. The citations were mostly undeserved and were given for minor insignificant reasons, but the fines were paid immediately and the citations not protested against.

Although a better control of the variables may be necessary, it seems unlikely that good drivers suddenly become bad drivers, so the hypothesis of police bias is confirmed. Although possibly caused the recent riots, such a bias is unacceptable. Due to the many citations, no bumper stickers are used anymore.

Table V -- Conventional Superstructure