Discourse Analysis: Its Development and Application to the Structure of News

by Teun A. van Dijk

Concepts and contributions of the study of language to the explication of the structure and meaning of texts are reviewed and applied to the study of news.

Developments in the last decade within such areas as text linguistics and, more generally, within the growing interdisciplinary study of discourse, have potential applications for the systematic analysis of mass media messages. Discourse analysis can make more explicit the classical approaches to "content analysis." It can also stimulate a research paradigm within mass communication that sees textual analysis not only as a method of research—for example, in the study of media effects—but also as an autonomous endeavor toward the construction of a sound theory of media discourse.

Because the study of discourse has become a large field in the past ten years, my discussion here must be limited to those aspects of discourse analysis that seem most relevant for the study of media discourse. Thus, I will pay little attention to those properties of discourse that can be characterized in terms of linguistic grammar in the strict sense, such as the syntax and semantics of (isolated) sentences. Rather, I will be concerned with more specific textual structures that have been neglected in linguistics. Similarly, I cannot go into the details of stylistic and rhetorical analysis of media discourse, although much work in this area still needs to be done. Finally, I will also limit my application to news discourse in the press, thereby neglecting TV, film, and radio discourse, the role of images in audiovisual forms of discourse, and other types of newspaper discourse such as advertisements and commentaries.

Teun A. van Dijk is Professor in the Department of General Literary Studies, Section of Discourse Studies, University of Amsterdam.
Of course, the study of discourse is not restricted to the structural analysis of texts. I will show that the psychology of discourse processing, concerned with the cognitive principles of the comprehension and remembering of texts, is of fundamental importance in mass communication research.

The 1970s saw a wide interest in the study of "texts" or "discourses" in such disciplines as linguistics, semiotics, literary scholarship, psychology, sociology, and anthropology.

In each of these disciplines, an interest in texts seemed to mark a paradigm shift with respect to earlier studies of the structures and functions of language. Besides a focus on the "system" of language, explicitly accounted for within structural and generative-transformational grammars [62], a new emphasis was placed on the analysis of language use (e.g., within cognitive and sociocultural contexts), on language use as social action (e.g., within the study of so-called speech acts [821]), and on the analysis of "natural" data (e.g., everyday conversations). In several of these areas, attention shifted from the study of individual words, phrases, or sentences to an analysis of structures and functions of actual forms of language use, that is, to discourse.

This wide interdisciplinary interest in the study of discourse has earlier historical antecedents, of course. Classical rhetoric, from the work of Aristotle to the present day, has always been concerned with the (persuasive) properties of discourse, and the sophistication of its analysis of rhetorical operations, such as the so-called "figures of style," has met with some competition only with the advent of structuralism (49, 58, 71).
Structuralism also brought a decisive reorientation to the study of literary discourse (11, 47) in such areas as classical poetics and literary scholarship.

This "structuralist revolution" in the classical study of discourse in the twentieth century has two main sources. First, structural linguistics provided the necessary methodological renewal by offering an explicit definition of structural units and categories and the formulation of rules. Second, anthropology, itself inspired by this development in linguistics, gave impetus to the very successful structural analysis of narrative (1, 4, 39, 72; for a survey, see 41). The systematic analysis of discourse would be unthinkable without these predecessors of the 1960s.

Yet, in addition, a more general account of language use, interaction, and communication is needed. Despite its earlier emphasis on abstract sentence grammars, linguistics has been at the center of this development, providing the necessary explicit methods for the systematic analysis of discourse. Besides the mainstream paradigm of Chomskyan generative-transformational grammars, the 1960s saw the formation of several "schools" of linguistic discourse analysis (3, 12, 24). First, the tagmemic school, centered around the work of Pike (70) in describing non-Western languages, has always paid attention to discourse, especially to stories and the structure of paragraphs (40, 61). Second, among the European approaches to discourse in linguistics, the systemic grammar of Halliday has inspired many studies (10, 27, 44, 59) at the boundaries of grammar and stylistics and now manifests itself mainly in the school of discourse analysis at the University of Birmingham (8, 9, 83). The third major influence came from those working in (mainly German) text linguistics and text grammar, who advocated the construction of grammars that would account for linguistic structures beyond the boundaries of the sentence (3, 12, 14, 67, 68, 69).

The linguists also had company in other fields. Sociolinguistics urged that more attention be paid to actual language use and that the hitherto silently presupposed social nature of language be taken seriously. Apart from dialectal and sociolectal variations in language use, or the study of the interdependence of linguistic forms and social categories such as situation, institution, age, gender, status, and role or group membership, sociolinguistics also was confronted with language use in discourse, especially under the impetus of work by Labov (54, 55). It has been increasingly accepted that language systems and language use are not autonomous, but are inextricably related to the interactional functions and the social contexts of verbal communication: language and discourse forms thus mark or "indicate" their relevant social parameters (81) and are treated as manifestations of social action of a specific kind.

Closely related to this development in sociolinguistics has been a growing interest in the study of many discourse forms (after the earlier study of myths and folktales) in their cultural context (2, 42) within the fields of anthropology and ethnography. Such studies typically show that
storytelling not only has different structural categories in different cultures, but also puts specific constraints on who can tell what to whom under what circumstances or, similarly, on how greeting rituals or other speech events take place in such cultures.

The emphasis on naturally occurring speech has led many to the analysis of everyday conversation. Inspired by earlier work in microsociology, mainly within the so-called "ethnomethodological" tradition (33), both linguists and sociologists have formulated a number of basic principles of the "dialogical" and interactional aspect of language use and discourse, such as turn-taking, strategic moves, and everyday storytelling and arguing (80, 85, 88).

Finally, following work on language comprehension inspired by the sentence-grammatical paradigm of generative-transformational grammar, psychologists, too, discovered discourse. Both cognitive psychologists and scholars from the new burgeoning discipline of Artificial Intelligence (AI) developed models for the processing of discourse and its representation in memory, which have been widely applied, for example, in educational psychology (e.g., in the study of reading). These psychologists not only formulated the important cognitive dimension of discourse interpretation, but they also developed their own structural models, such as narrative grammars, and were the first to work out explicitly the now well-known assumption that understanding discourse presupposes vast amounts of general knowledge of the world.

Before bringing the theoretical and analytical results of these various approaches to discourse to bear on the study of masa communication, I will offer a brief summary of some basic principles for the analysis of discourse.

Since the diversity of theoretical, methodological, and terminological persuasions is too impressive to allow a short synthesis, I will here formulate the main properties of discourse against the background of my own earlier work (12, 14, 16, 18). I refer to other work for details or different approaches.

Verbal utterances, such as sentences, discourses, texts, or messages (I will henceforth use the term "discourse"), are usually analyzed first on different levels. The structures at each of these levels are accounted for by specific sub-theories or even sub-disciplines of linguistics. Thus, there is phonology, accounting for the structure of sounds and intonation, morphology, formulating the principles of word formation, syntax, providing the rules according to which words of different categories can be combined into grammatical sentences, and semantics, dealing with the meaning of words, phrases, sentences, or whole discourses by formulating tales of interpretation. As opposed to such "underlying" meaning structures, the phonological, morphological, and syntactic
expressions manifesting this meaning are sometimes simply called "surface structures." In practice, much of the work in discourse analysis has concentrated on semantic structures, that is, on meaning, because earlier work on sentence grammars tended to focus on surface structures.

In addition to the levels, different units of analysis can be distinguished in discourse: individual words (lexical items), various structures of the clause, whole sentences, sequences of sentences (paragraphs), or whole discourses. The overall topic or theme of a discourse, for instance, can be studied only at the semantic level of the discourse as a whole, not at the level of individual words or sentences. Hence, a rather rough distinction is usually made between 'local' and 'global' structures of discourse, with the former pertaining to sentences and immediate sentence connections and the latter to larger segments of the discourse or the discourse as a whole.

Next, cutting through the various levels mentioned above, and both locally and globally, different dimensions of analysis can be distinguished. Thus, stylistic variation can occur at several levels, such as lexical choice, word formation, or syntactic structures. Similarly, rhetorical operations (such as alliteration, parallelism, metaphor, or irony) also require definition on various levels. Finally, there are different modes of the manifestation and use of discourse, such as spoken or written/printed discourse, monologues, and dialogues. The various units, categories, dimensions, and levels, along with the rules defining them, will all be called "textual." However, discourses are not just isolated linguistic "objects," but are integral parts of communicative acts in some sociocultural situation, which I will call "context." Thus, it is a contextual property of the discourse type "verdict" or "plea" that it is rightfully used only in the courtroom and by a judge or lawyer. At the boundary of text and context, the pragmatic analysis of discourse is concerned with the dimension of action in which a discourse is taken as some conventional forro of social action (promise, threat, question, congratulation), called a "speech act."

I have provided, in extremely succinct terms, some elementary notions of discourse analysis. The various schools of discourse analysis mentioned above can be distinguished, in part, on the basis of their specific interest in some textual or contextual property. Thus, some people will exclusively study discourse style, or intonation in spoken discourse, or overall meaning, or specific social constraints on the context. Similarly, there can also be specialization in certain discourse types or genres, such as everyday conversation, stories, classroom discourse, textbooks, proverbs, or news. Each discourse type, then, could—or rather, should—be characterized in terms of a specific combination of various textual and contextual properties. A judge's verdict, for instance, should have a specific (formal) style and is constrained to specific overall meanings (themes, topics).
Thus, to focus on news discourse (as I will below) requires a full analysis of its various levels, units, dimensions, modes, and social contexts. Of course, a relevant analysis would focus on those structural aspects that are typical. Structural differences exist not only between two such disparate discourse types as an everyday conversation and a psychology textbook, but also between a spontaneous everyday talk and a job interview, although both are dialogues. It is therefore not easy to specify in general what the properties of a discourse are at the various levels and for the respective units and dimensions. Nevertheless, we can specify some fairly general characteristics, which then can be further detailed for news discourse.

1. **Functionality.** If a discourse is taken to be the utterance of a sequence of sentences in some social context, then the various properties of such a discourse are assumed to be functional with respect to various aspects of the social context. That is, both surface structures and meanings are produced and understood as indications about characteristics of the speaker (e.g., intentions, wishes, moods), the relations between speaker and hearer (e.g., confidence, intimacy, power), and the type of social situation (e.g., a court trial, a school lesson, a birthday party). This will hold for surface structure style, such as lexical choices and sentence structures, and also for the possible topics or themes talked about or the speech acts that may or should be performed with the utterance of the discourse. The functionality also holds, therefore, "within" the discourse: the surface structure not only expresses or indicates social structure, but also, and even primarily, is meant to express underlying meaning (35).

2. **Meaningfulness.** A textual sequence of sentences is distinct from an arbitrary collection of sentences in the sense that, in principle, such a sequence should be meaningful. One of the typical conditions for meaningfulness of a discourse is some kind of unity, which is usually described in terms of local or global coherence. Local coherence means that subsequent clauses and sentences are meaningfully related, because the facts to which they refer are causally related or because the propositions expressed by these clauses or sentences are related (one proposition may be a specification, generalization, or example of a previous proposition). Global coherence pertains to larger parts of the discourse; this kind of global unity is usually described in terms of such notions as "topic" or "theme." Such themes or topics are accounted for theoretically in terms of so-called "semantic macrostructures." Thus, a fragment of a discourse or a whole discourse is considered to be globally coherent if a topic (represented by a macroproposition) can be derived from such a fragment. Note that part of the meaningfulness criterion for discourse is not only that (sequences of) sentences have meaning, but also that they are "about" something: they refer to (real or imagined)
Discourse Analysis and the Structure of News

course is less studied as a form of “social practice” in its own right, for which it is a legitimate aim to make explicit the inherent structures at all levels of analysis.

Second, a discourse analysis primarily aims at the explication of qualitative data rather than quantitative data. Of course, quantitative measures may well be based on an explicit analysis of a more qualitative kind.

Third, while content analysis is primarily based on observable, countable data, such as words, phrases, sentences, or stylistic features, a discourse analysis will—apart from making explicit such surface structures in terms of modern grammars—also pay attention to underlying semantic structures and make explicit implications, presuppositions, connections, strategies, etc., which usually remain implicit in the discourse. It will try, in terms of empirical theories, to find the rules or principles underlying the structures, the production, and the comprehension of media messages.

Finally, a discourse analysis will be part of a more embracing cognitive and social theory about the rules and strategies that underlie the production and understanding of media discourse. Instead of merely correlating, it will try to explain, in precise cognitive models, how various structures of media discourse come about and how media discourse is understood and represented in memory. Hence, the relation between content properties and specific “effects” is split up in terms of a number of highly complex cognitive and social-psychological models of information processing.

Of course, these distinctive features of discourse analysis are taken to be relative to content analysis as a whole. There are certainly studies that have come close to one or more of these aims. Many of these studies have appeared in England (and in Germany; see 84) and often explicitly mention their opposition to traditional communications research in the U.S. Work done at the Centre for Contemporary Cultural Studies of the University of Birmingham (e.g., 43), influenced by several theorists in France (such as Barthes, Althusser, Poulantzas, Lacan, and Pécheux), has especially focused on the ideology of media discourse. Through more explicit linguistic discourse analysis (and another theory about the role of the media in society), such work attempts to uncover implied meanings that represent ideological positions. A similar goal can be found in the media analyses of some linguists (e.g., 28, 53) who, by meticulous syntactic analysis, show that the very sentence structures in news can mask who are the responsible “agents” in some events and that newspapers (e.g., the Sun and the Morning Star) thus can be differentiated linguistically according to their underlying ideologies. Finally, the Glasgow University Media Group (36, 37), in its well-known Bad News studies of television news, also applies a systematic verbal and visual analysis, uncovering how the very photographs or noun phrases employed can bias the news, for example, in favor of employers and against
(striking) workers. These few studies indicate the kind of approach I hope to stimulate and, if possible, to make even more systematic and explicit (for detail, see 20).

The theses formulated above can be illustrated and further elaborated for a specific kind of media discourse—news in the daily press.

I will focus here on textual structures of news and largely ignore the various contextual conditions and constraints on such textual structures. The latter have received much attention in recent years, especially in sociology (7, 26, 32, 38, 76, 87). These studies emphasize that news is not simply an (incomplete) description of the facts, but a specific kind of (re)construction of reality according to the norms and values of some society. They also show that news production is part of a complex of professional routines for the management of possible sources, the interaction among journalists, and the possible "formulations" of reality.

To such a social background, which will not be discussed further here, should be added a cognitive perspective. The perception, construction, and formulation of reality as news events not only underlie social routines of journalistic (inter)action, but also fundamental strategies of information processing. I stressed above that media discourse should not be seen merely as a ready "product" of news-gathering activities, but as the manifestation of a complex process in which knowledge, beliefs, and opinions are matched with existing or incoming information about events, the social contexts of news production, and representations of the reading public. More specifically, it should be stressed that news production is not a direct representation (biased or not) of events, but rather some form of discourse processing (19). Reporters will seldom be direct witnesses to events; rather, their data are mostly other discourses, such as eyewitness reports (60), press conferences, press releases, statements of officials, interviews, documents, or news of other media and press agencies. Hence, the construction of news is most of all a reconstruction of available discourses. The interpretation processes, the representation and the retrieval from memory of these discourses—at various stages until their final form—make up one of the basic but ill-understood components of news production.

Work in the psychology of discourse processing has contributed much to the understanding of these processes (22, 29, 30, 48, 51, 78). This work shows what decoding operations and strategies are at work in short-term memory during the understanding of discourse, how discourses are represented in and retrieved from long-term memory, and what the role is of world knowledge in these processes of understanding and representation. It has shown, for instance, that memory for discourse is especially semantic and that, above all, the themes or topics—that is, the semantic macrostructures—can be retrieved from memory, depending also on the schematic organization of the discourse, on the one hand,
Studies have also offered insight into the roles played by knowledge of and beliefs about the world in the processes of understanding and producing discourse (79).

These results are fundamental for a sound theory of the production and understanding of news as forros of discourse processing. It is now known more or less explicitly, in precise cognitive models, what it means to be “biased,” what the relation is between textual structures and representations of "reality," what the precise conditions are of the well-known “selective perception” of both journalists and readers. Similarly, the criteria of newsworthiness formulated by Galtung and Ruge (31) can be reconstructed in even more precise cognitive terms.

This psychological work not only provides a basis for the sociological work done on news production as “social practice”—and thereby an account of the structures of the roles of underlying ideologies—but, above all, insight into the very structures of news discourse. The organization of news discourse is both a result of and a condition for the cognitive operations of journalists and readers, respectively, in the production, reproduction, or understanding of the news "data." Besides the social routines, then, the cognitive routines that lead to the conventional structuring of news discourse must be formulated. I will focus here on the textual "consequences" or "conditions" of these basic cognitive principles.

Within the general framework sketched above, I will review some of the properties of news discourse, touching on cognitive expectations, stylistic structures, local and global coherence, and schematic superstructures. A full-scale analysis of even one news article would be very complex, depending on the degree of formality desired, so my analysis must be fragmentary and merely illustrative. In a description of a large corpus, particularly, selective analysis of some relevant features is practically always necessary. As was indicated above, the analysis will not only be purely "structural," but will also involve some cognitive implications. This is necessary because the actual "understanding" of a news discourse depends not only on its manifest structures, but also on already presupposed cognitive information and processes (strategies) of interpretation and representation. In other words, empirically speaking, the structure of a news discourse is ultimately the one assigned to the text by the reader. General structural features, then, are abstractions from such cognitive representations.

An important cognitive aspect of news discourse interpretation is the set of expectations the reader brings to bear even before reading the discourse itself.

Already established in the reader’s cognitive control system, regulating the flow of information between short-term memory and long-term memory (knowledge and beliefs), will be some overall goal for reading
the discourse, such as the acquisition of knowledge about political events. The reader will also have activated and stored in the control system the relevant general knowledge about the newspaper and its dominant views on political matters (ideology) and possibly will have concrete expectations about the type of semantic information (e.g., foreign affairs) on some specific page. Episodic memory (the part of long-term memory that stores all incoming experiences) will yield, by a strategy of "reminding," information about events that was gained in the last few days. Finally, the news reader will have activated general knowledge about the structural format, e.g., the so-called news discourse "schema" (see below), of news articles in the particular newspaper.

All these types of information are crucial for the structural or cognitive analysis of a particular article. For instance, they allow possible disambiguation of short headlines or the activation of relevant knowledge and beliefs about some ongoing political event, such as general knowledge about war when reading of, say, the Israeli attack on Lebanon in June of 1982. Such packets of organized conventional knowledge, called "frames" or "scripts" (79), are important determinants of the inferences necessary to understand each word, sentence, or sentence connection. In addition, such scripts may pertain specifically to the political ideologies of the parties involved in international conflicts.

A stylistic analysis of surface structures—the study of variation in the expression of underlying meanings and reference of utterances—can reveal both cognitive and social functions of language choice.

The grammatical choices made for a discourse (or by a speaker)—the choice of specific lexical items over others, or the expression of underlying propositions in different syntactic structures—may sometimes be arbitrary (and beyond cognitive control), but may also be functional. One such type of variation can involve the cognitive context of reading. For example, simple, short sentences, with standard noun-phrase verb-phrase (or subject-predicate-object) structures, may be easier to understand than long, complex sentences.

More interesting, though, are the possible social implications of stylistic variation. First, a style register can indicate some specific

---

1 For detailed analysis of these strategies of comprehension, using a Newsweek text, see (22).

2 For details of a computer program simulating this kind of "subjective information," see (5).

3 A complete description of the surface structures of news discourse, that is, of the precise graphical and especially the syntactic and lexical or morphological properties of each sentence, is of course possible, but not always relevant. For this kind of analysis, which would rather be interesting for linguists, the reader is referred to the usual linguistic grammars of sentence structure (for an introduction, see 62).
Discourse Analysis and the Structure of News

degree of formality or familiarity. Second, it can indicate the specific beliefs, opinions, and ideologies of the journalist/newspaper that is describing the events. Of particular interest is the choice of lexical items used to denote some participant in the events described.

Thus, Halloran (45, p. 95) shows that the descriptions used in the press to characterize participants in demonstrations against intervention in Vietnam ("hooligans," "thugs," "mob," "horde," etc.) will often have negative implications ("connotations"), and the same holds for the actions and properties described by verbs and adjectives. Similarly, in a case study I conducted in Amsterdam about the coverage in the national and local press of "riots" following a police action against squatters, I also found systematic differences between the lexical description of the demonstrators by the national press agency and by the popular (conservative) newspapers, using the Dutch equivalents of "hooligans," "rioters," "rowdies," and "thugs." Also, the lexical items used to describe the police and their actions, although they were also violent, were usually more neutral (for details, see 21).

The Glasgow Media Group (37) has shown that, in the coverage of industrial affairs, the press will typically choose lexical items to denote workers and their actions ("strike," "disruptive actions," etc.) that are more negative than the items chosen to denote the actions of the employers. Also, the workers are typically described as "demanding" and the employers as "offering," and not the converse. Thus, even at this restricted level of stylistic variation in lexicalization, basic opinions and ideologies about social participants and social actions can be expressed.

Fowler et al. (28) have focused in particular on the syntactic structures of sentences. They have discovered that if, for instance, the police are reponed to be the agents of violent actions, such agency is not expressed in the more active "first position subject" position, as is usual (see 35), but rather "suppressed" in passive sentences and nominalizations ("many demonstrators were hurt").

Although lexical choice is a typical phenomenon at the local level of sentence, it is also relevant for discourse analysis because a whole discourse will typically exhibit some kind of stylistic coherence: throughout the text, lexical choices will be made from the same register and, in order to denote the same referents, will be subject to the same evaluative dimensions. Even "neutral" words, such as "sympathizer," can receive a negative connotation in certain contexts. In the process of macrostructure formation—that is, the derivation of overall topics or themes (see below)—this means that an overall negative concept of a participant is formed, even if not all denominations are actually negative. And it is the overall concept that is most readily retrieved in memory (51).

* In the 1970s, many studies paid attention to the ideological bias of lexical choice in the press (e.g., 6). For further aspects of stylistic analysis, see (10).
Although lexicalization is usually treated as a surface structure phenomenon of language use, it is in fact halfway between the surface structure and underlying semantic structures of meaning. The use of “hooligan” vs. “demonstrator” is not just an equivalent alternative choice to express the same underlying meaning. There is, of course, a difference in meaning, but only in terms of how identity is referred to: the same participant is denoted by the two expressions. The same holds for the lexicalization of predicates expressed in verbs and adjectives. The combination of such concepts in propositions can result in evaluative implications for whole propositions (see 63, 64).

In the scope of analysis beyond the sentence level, news discourse, like all discourse, should exhibit local semantic coherence, which, like other meaning aspects, should be expressed in surface structure (see 44). As shown above, local coherence of discourse is defined in terms of the relations of meaning or references that exist between propositions: locally coherent propositions will denote “related” facts in some episode or will themselves be functionally organized (e.g., by a “general-particular” relation). Unlike simple stories, in which conditional relations such as those of cause or reason and consequence may dominate (see 13, 17), news stories will also have functional or rhetorical relations of local coherence (see 18, 40, 75).

The most pervasive news discourse type of this kind of local coherence appears to be specification. Some fact is described, and a next sentence will give details or particulars, thus organizing the discourse along a dimension of “general toward particular” (a dimension that will be further discussed below). Thus, The Guardian reports on July 8, 1982, on page 6, under the headline “Soviet Embassy compound hit in Beirut duel,” as follows:

Beirut: Israeli forces and beleaguered guerrillas blasted each other in a 20-hour duel in and around Beirut and police reported 22 people killed and 39 wounded before the guns fell silent at midday yesterday. There was no breakdown of civilian and military casualties.

The first few propositions of these two sentences are properly connected by conditional/temporal relations (among the facts denoted): mutual fight causes casualties, which condition the information reported by the police. The information that some detail was lacking in the police report is not “conditioned” by the previous facts, however. Rather, it is a specification of the previous proposition. At the same time—and with important cognitive implications—this sentence presupposes that such information (about the breakdown into civilian and military casualties) could be expected, or normal, or a property of previous reports about the Israel-Lebanon war (as indeed is the case—a bit of information stored in the reader’s episodic situation model of this war that has been construct-
Discourse Analysis and the Structure of News

ed through previous news items). With respect to this implied knowledge, the last sentence also functions as contrast or denial, which means that, in terms of local coherence connections, propositions can have several functions at the same time. "Israel used tanks, gunboats, and field artillery" is a typical example of specification.

The inverse operation can also hold. The next paragraph in this article begins,

Yesterday brought the total of officially reported Palestinian and Lebanese casualties in Beirut to 2,633 killed and 3,612 injured since the invasion of June 6.

From the "local" action of today, a generalization or summarization is made. Again, this is not a description of the actual events as such but a characterization at a higher level, an inference, about the whole event by the journalist. Note that the definite use of the noun phrase "the invasion" presupposes (or by presupposition reminds the reader of) the major initial action of this war, thus reinstating from memory the relevant episodic knowledge. Below, I will show that "previous events" is a special conventional category in the overall organization of news discourse. I will therefore provisionally assume that news discourse has both conditional and functional coherence relations among propositions as soon as the events themselves are described. In that case, the functional relation will tend to be specification, that is, the description of lower-level particulars or details. Generalization will occur as soon as a comparison or summary is necessary with respect to previous events. Contrast or denial will usually come into play when information is compared to expectations derived from previous events or from memory scripts (and ideology). Thus, Halloran (45) observed that peaceful demonstrations can be described with such denials of violence, a routine strategy that may lead readers to associate demonstrations (or even the reportedly peaceful demonstration) with violence anyway.

To be meaningful, a discourse should not only be locally coherent, but also globally coherent—there must be some kind of "semantic unity" to the whole discourse.

Intuitively, this semantic unity is obtained by assigning some theme or topic to the discourse or to a fragment of the discourse. In the case of the newspaper article just analyzed at the local level, this means that the overall meaning or theme is "Israel and guerrillas (PLO) blasted each other, yesterday. Such a theme, according to my theory of discourse (14, 16), is defined by a macroproposition and not just by isolated concepts (e.g., "fight"). Such a macroproposition is derived from the information, represented in the respective propositions expressed by the text, of the
discourse as a whole. This derivation takes place through macrorules, such as deletion (of irrelevant detail), generalization, and "construction" (in which component actions, for instance, define an overall action).

For the news article on the Lebanon war, this is indeed the case. The rest of the text provides details about the weapons used, the damages to the Russian Embassy (forming the other macroproposition of this discourse), and other damages. In order to derive the relevant macropropositions, the reader needs not only the information expressed in the discourse, but also more general knowledge of the world. The reader must know that tanks and gunboats are weapons used in wars and that this "use" is against another party, that is, the "enemy." And, if several buildings of the Russian Embassy are reported to be damaged "during" the Israeli shelling, knowledge of the world (the "war" script) tells the reader that probably the damage was caused by the shelling. Finally, the overall theme, defined by one or more macropropositions thus derived, can also involve overall opinions—evaluations of the facts—derived from the way the facts are described locally, as in the news coverage of the squatter "riot" in Amsterdam.

Besides the intersubjective macrostructure as indicated by the text, I should emphasize the role of the subjective macrostructure as assigned by the reader, according to the cognitive model. In the latter case, the macrostructure can be personal and subjective, bringing to bear personal knowledge, beliefs, and opinions, and thereby defining what is important or relevant for the reader. From the local information "A and B blasted each other," the reader who disapproves of the Israeli invasion will construct the macroproposition "the Israelis (again) attacked the PLO/West Beirut" or even "the Israelis again violated the cease-fire." Macrostructures have been shown not only to organize information from discourse at a rather high level in memory, but also to be recalled much more readily than local micropropositions (50, 51). The (subjective) macrostructure will be the essential information called on in reading the news for storage, recall, and further use in constructing a "picture of the situation" and, eventually, knowledge of and opinions and attitudes about the world (for radio news, see 57).

The overall organization of news discourse reflects this importance of macrostructures. These will typically be expressed by titles or headlines, by initial or final summaries, or by leads. Macrostructures are usually signalled by (a) a prominent position in layout, (b) a change in typeface, and/or (c) bold or capital letters. The headline, thus, will typically express the most important macroproposition, where "importance" is defined in terms of general knowledge and beliefs defining the newsworthiness criteria (19, 31). The lead, often printed in bold type (although not in the Guardian example cited here), will express, in a first few sentences (which are, by definition, "thematical sentences"), the full macrostructure of the news discourse. Following sentences will then progressively specify further details of the events, with the less impor-
Discourse Analysis and the Structure of News

Important ones at the end (with the practical consequence that these can, if necessary, be cut by the editor). Unlike argumentatively structured discourse, such as scholarly papers, where the important conclusion comes at the end, and unlike weekly news articles, which may express an opinion at the end, news in the daily press is organized by the principle of relevance or importance, along a dimension of decreasing prominence with respect to the macrostructure. This means that one can read only the headlines or the lead, or only some part of the discourse, and still process the most important information. In fact, as Reder and Anderson (73, 74) have shown, readers (at least of textbooks) will often remember no more of texts than their summaries (see also 51).

For the processing of news discourse, the headlines, the lead, or the first thematical sentences are crucial for various reasons: they organize ("attract") attention for specific articles; they allow one to decide whether or not to read the rest of the discourse; they give the main theme, even without further reading; they activate the relevant knowledge from memory that is needed in order to understand the rest of the text; and, last but not least, they form a macrostructure that will serve as an important strategic cue in controlling the local understanding of the subsequent text. Indeed, a lack of title, or a biasing title, can inhibit or fully distort the local comprehension of a text (see 23, 52).

From a methodological point of view, the importance of macrostructural analysis in mass media research is that it allows the explicit definition of main topics or themes in messages, even for those cases where these macrostructures are not specifically expressed in surface structures, that is, in titles, leads, or thematical words and sentences. More specifically, such an analysis allows the demonstration of cases in which a headline does not express the macrostructure of a text, a phenomenon often observed in the case of a "misleading" headline. In such a case, the newspaper/journalist has, so to speak, "upgraded" some detail to macrostructure status, e.g., on the basis of subjective criteria of relevance. Given the important functions of macrostructures in comprehension, representation, and hence in recall, a sound theory and derived practical method of analysis is a powerful instrument for mass communication research and a renewal for classical content analysis.

The various functions of propositions in a discourse can become conventionalized in a given culture and for given discourse types in such a way that they can be analyzed as constitutive structural categories for that discourse type.

Thus, the initial sentences of a story can globally function as the setting of a story. In the same way, a story can have a complication, a resolution, and an evaluation (see 12, 56). Similarly, since the first classical theories of the syllogism, argumentative discourse has been
analyzed in terms of different kinds of premises and a conclusion. All these categories of a discourse type together form a conventional schema, which I have called a superstructure (15, 16). This schema has a hierarchical nature: some categories should be defined at higher levels than others. Theoretically, a schema or superstructure is a functional organization of the macrostructure of a discourse. In other words, a superstructure is the overall "form" for the overall content defined by the macrostructure. And indeed, such forms, which are some kind of higher-level textual "syntax," can be defined explicitly in rule systems or grammars (though such grammars are of course different from the usual sentence grammars). Narrative grammars have been developed, then, especially for stories (see 17, 41). Of course, such conventional categories are particularly relevant for those discourse types that occur frequently in some culture and for which the ordering and other constraints on the semantic content have become conventionalized and hence learned.

Since many participants in Western culture are regularly confronted with news discourse in the press (or on television), such news articles perhaps also can be assigned a conventional superstructure. Such a superstructure is postulated in Figure 1.

The categories of "headlines" and "lead" in newspaper discourse are well known and need no further explication. Note that both together define a higher-level category of "summary" and "introduction," with the semantic constraint that this "summary" must express the semantic macrostructure of the text as a whole; at the same time, it functions as the "introduction" to the text, because here the main events and participants and the location and time are introduced (such that the body of the text may, even pronominally, co-refer to these).

Categories 1 ("summary") and 2 ("episode[s]") together form the proper "news story" category at a higher level. It is with respect to this higher-level "news story" that the "comments" are given. The "comments" are the typical expression, signalled as such, of the beliefs of the journalist/newspaper. An "episode" can be described first in terms of "reminding" the reader of information given earlier (or elsewhere) in the newspaper. "Antecedents" provide the information about the facts that precede the actual events (and that may or may not have been reported before). "Actual events" is the kernel of the news story and describes the main (new) events now in focus. Usually, though, especially in "quality" newspapers, there will be a section, which may run from one sentence to a whole paragraph, with further "explanation" of the actual events, that is, some specific "context" of the actual events (e.g., a demonstration in Europe in the context of the visit of the U.S. president), as well as "background," which provides the historical, cultural, or political information about the actual events, participants, countries, or social problems.
Discourse Analysis and the Structure of News

Figure 1: A proposed conventional superstructure of news discourse

1. Summary/introduction
   1. 1. Headlines (with super-, main-, and sub-headlines, and captions)
   1. 2. Lead
2. Episode(s)
   2. 1. Events
      2. 1. 1. Previous information
      2. 1. 2. Antecedents
      2. 1. 3. Actual events
      2. 1. 4. Explanation
         2. 1. 4. 1. Context
         2. 1. 4. 2. Background
   2. 2. Consequences/reactions
      2. 2. 1. Events
      2. 2. 2. Speech acts
3. Comments
   3. 1. Expectations
   3. 2. Evaluation

These categories are theoretical; they define a prototypical news discourse. This means that, in an actual news discourse, some of them need not occur explicitly, e.g., because it can be assumed that readers already have that information or because such information is deemed to be irrelevant. In the latter case, a proper understanding of the description of the actual events can become partial or even biased if readers do not, in fact, have the information.

The hierarchical schema also defines the ordering of the text as a whole, from left to right (or from top to bottom). However, the superstructure may undergo transformations. For instance, category 2.2, giving consequent events or speech acts (typically, statements of politically important participants or observers), can be placed earlier and "previous information" or "antecedents" (categories 2.1.1 and 2.1.2) placed later in the text.

The categories enumerated above can be specifically defined by a rule system (see 62) or by a hierarchical schema, as in Figure 2. The terminal categories of this schema are filled with (macro)propositions of the newspaper discourse. I have applied the schema to information from another newspaper article from the same page of The Guardian, headlined "Kremlin talks tougher," in which the negative reaction of the Russians is described with respect to U.S. plans to evacuate the PLO from West Beirut.

Although with such a short text, not all categories are filled, it is easy to find news texts in which these categories can be discerned. Further empirical work should be done in order to test the relevance of this news schema. It should also be investigated experimentally whether such a news schema has some cognitive "reality": do readers in fact organize a news discourse according to such categories; do they "expect" them or
Kremlin talks tougher

By Mella Piek

The Soviet Union, which has been reluctant to respond to PLO pressure to become more directly involved in the Lebanon crisis, yesterday hinted that it might adopt a tougher stance if President Reagan's plan to include US troops in a multinational peace force goes ahead.

Already angered by Israeli shelling of the Soviet Embassy compound in West Beirut, Moscow's first public reaction to President Reagan's plan was to accuse the US of "preparing for direct military intervention in Lebanon." Radio Moscow also said that President Reagan would be acting illegally if US landing craft were on their way to Lebanon.

The Kremlin, which has accused the US of supporting and, indeed, encouraging the Israeli invasion of Lebanon, is known to be concerned about the possibility of a US troop presence there, however temporary. Analysts have been arguing that one important reason why the Soviet Union has been acting with restraint so far in the crisis has been the Kremlin's desire to do nothing that might give the US the justification for direct military intervention.

The Kremlin is obviously waiting to see what becomes of the Reagan proposals. Meanwhile, Tass yesterday sharply condemned the attack on its embassy buildings on Tuesday night, which caused extensive damage to the six storey Soviet trade mission in Beirut, and a nearby apartment block for Soviet staff. Tass claimed that Israeli artillery had directed heavy fire on the embassy areas.

The official Soviet news agency stopped short of accusing the US of collusion. But the Soviet Foreign Minister, Mr Gromyko, again accused the US of the direct encouragement of Israeli aggression.
Note: Terminal categories in the schema are propositions. Those in "headline" and "lead" are macropropositions. The main event is the Soviet Union's tough reaction against Reagan's plans. Hence "antecedent" events are Reagan's plans. The context for the Soviet tough reactions is (its anger about) Israeli shelling of the Soviet embassy. But this is an "antecedent," namely, for $p_1$: Soviet Union condemns Israel. The "background" explains why the Soviet Union is tough now and not before, and provides the (earlier) Soviet interpretation of the invasion. Under the terminal categories are some summarizing (macro) propositions. Part of "antecedent" could be "previous information" (the Embassy anide on same page).
"miss" them when they are not given? Comparing different narrative schemata for news stories, Thorndyke (86) arrived at a negative conclusion, but other research on the role of schemata in psychology points in another direction (51, 65, 66, 77).

A final kind of structure, provisionally called "presentational," can be mentioned here. Although these structures could be ranged under the category of surface structures, I treat them separately because they are relevant at all levels and thus constitute a separate dimension. First are such obvious properties, well known from practical insights into news in the daily press, as page number, position on the page, size, printing types used, and number of columns. Other aspects of presentational structure are the presence of photographs and, in television news, the relations between spoken text and stills or film.

In fact, some of these structural properties could be called rhetorical and can be compared with the "performance" or "delivery" of spoken discourse. Nonverbal information, in this case, is provided by other semiotic means, such as location and size. Obviously, these semiotic signals of various kinds have, or at least are intended to have, specific cognitive implications: location, size, etc., correlate with importance or relevance of the news and therefore can be considered as macrostructure signals at an intertextual level. That is, they signal which of many facts or events is relatively more important. Empirical research about newspaper reading, gaze, skimming, etc., will reveal the relevant implications of these various presentational structures in comprehension and recall. Findahl and Höijer (25) have shown in a number of experiments the relevance of presentation of TV news on recall (the role of visuals and also of reformulations).

After a brief summary of some major directions in discourse analysis and of some principles of discourse structures in general, I have tried to show that news discourse in the daily press can be described systematically and explicitly on the basis of results of text linguistics and discourse analysis in the last decade. I have shown that surface structure (stylistic) analysis, as well as the analysis of local and global coherence (macrostructures) and that of a specific news schema, may be relevant in an account of mass media "messages." Such an account should be embedded in a cognitive model of news production, comprehension, and recall.

REFERENCES

Discourse Analysis and the Structure of News

Discourse Analysis and the Structure of News


