

---

TEUN A. VAN DIJK

## DIALOGUE AND COGNITION

### 1. PROBLEMS AND AIMS

1.1. Dialogues are verbal interaction sequences performed among language users. In order to be able to adequately accomplish their respective actions which constitute a dialogue, these language users must 'go through' a number of highly complex cognitive processes. It is the aim of this paper to briefly discuss some of the properties of the processes and representations involved in the cognitive management of dialogues.

The theoretical background of our discussion is the current work being done in the theory of discourse, and in particular the experimental research in cognitive psychology on processes of discourse production and comprehension. At the moment this kind of research is also done, from a theoretical point of view, in the domain of artificial intelligence. In the present paper we will have to see which additional properties must be postulated in a processing model for discourse in order to be able to account for specific aspects of dialogue. Our discussion will remain informal.

1.2. A cognitive analysis of dialogue indirectly involves many aspects of dialogues, e.g. structural properties on the one hand, and social and cultural conditions on the other hand (which also all need to be represented cognitively), which however cannot *be* discussed in the framework of one small paper. In order to stress that a complete cognitive model requires processes, operations, strategies, etc. which are linked to these structural and social properties of dialogues, we will nevertheless briefly list some of them:

1. phonetic and phonological relations between sentences, turns or moves of the verbal interaction sequence (intonation, stress, pitch, etc.); paraverbal activities;
2. morphological and lexical properties of sequencing (specific words as 'openers', 'closers', 'pass' indicators, pragmatic particles and other dialogue connectives)
3. syntactic properties and relations (incomplete sentences, completion of sentences by next speakers, phrasal boundaries of turns, repetitions, syntactic expressions of coherence phenomena, e.g. PRO-forms, etc.)

4. semantic relations, both intensional and extensional: local and global coherence (conditional relations between propositions expressed, establishment of topics of conversation, topic changes, etc.)
5. pragmatic properties of the dialogue, taken as a sequence of speech acts: local and global pragmatic coherence (relative appropriateness of speech acts, global pragmatic 'point', etc.); schematic structures;
6. interaction management of the dialogue taken as a sequence of 'moves' or 'turns' (turn taking and changing of turns, strategies and their rhetorical and stylistic devices, etc.)
7. socio-cultural properties of the interaction sequence, the participants and the context (categorization of actions and agents, conventions of social 'frames', institutional constraints, cultural rules, etc.).

Clearly, this list is not complete and also does not show the multiple relations between the various of analysis. The point is that language users in the dialogue are 'doing' many things at the same time, which requires cognitive planning, execution, monitoring, comprehension, storage and retrieval, etc. We will limit ourselves to only some of these cognitive operations, leaving the analysis of the further aspects to other approaches, e.g. those exemplified in the other contributions to this volume. We have already stressed above, however, that these other approaches, e.g. both the formal (game-theoretical or other) and the more empirical (e.g. the ethnomethodological analysis of conversation), may well have relevant results which may become components in a full-fledged cognitive model of dialogue.

## 2. THEORETICAL BACKGROUND

2.1. The theoretical background for our discussion features two large areas of research. The first is the general theory of discourse, including e.g. text grammars and other theories specifying various structures of discourse. The second is the psychological theory of discourse production and comprehension, which at some points is closely linked to the more linguistic' analysis of discourse structures.

2.2. The linguistic basis of a theory of discourse has been studied — e.g. under the label of 'text grammar' — first of all in the perspective of the extension of grammatical models towards structures beyond the sentence'. Apart from

---

rather obvious generalizations of rules holding for clauses within complex sentences towards sentence sequences in discourse, this approach allowed that the focus of e.g. semantic analysis also became directed on phenomena, such as 'coherence', which were neglected before. Taking text ' as the formal unit underlying a discourse utterance, it was also discovered that discourses should not only be analysed in terms of linearly or locally coherent sentence sequences, but also in terms of more global textual structures. Thus, the semantic characterization of themes' or of a discourse in terms of 'macrostructures' appeared to be an important task of such a 'global' analysis.

2.3. Besides the crucial cognitive implications of this kind of linguistic analyses of discourse, to which we will turn below, the various categories, rules and levels of this linguistic theory also appeared necessary for a treatment of other, 'non-linguistic', structures of discourse. Thus, theories of narrative and argumentation for example feature global 'syntactic' categories which require global semantic 'content', viz. propositions at a macrostructural level of meaning.

operations should not be limited to sentential structures alone, but also requires as their basis all kinds of sequential (inter-sentential) and more global textual units, rules and constraints.

In other words, the analysis of various kinds of discourse structures requires an integrated, interdisciplinary theory. It will be left open here in which respect the units (e.g. text') and rules involved belong to the domain of grammar or to a theory of language use and (other) communicative forms.

2A. A theory of discourse also involves a pragmatic component, in which (utterances of) sentences are interpreted as speech acts of which the appropriateness is determined by conditions formulated in terms of certain contextual properties of the communicative situation (knowledge, beliefs, wants, evaluations and certain social relations).

Relevant here, especially also for a theory of dialogue, is the fact that, pragmatically, discourse can be defined in terms of speech act sequences. Parallel to the local and global coherence at the semantic level, speech act sequences also should exhibit local and global coherence. Thus, a given speech act may be appropriate only relative to certain preceding or following speech acts, of which it may function as a condition or consequence :

- (1) Could you please tell me the time? I have no watch.

In this example the assertion accomplished by the utterance of the second sentence of this discourse may sometimes (e.g. for a stranger) only be appropriate, and 'relevant', when following as an explanation for a previous request.

Similarly, sequences of speech acts may be globally coherent due to the accomplishment of a 'macro-speech act', such as a global request of a letter, global assertion of a newspaper article or lecture, or threat of a ransom note.

2.5. It should be noted that much of the work very briefly referred to above did abstract from the distinction between *monological* and *dialogical* discourses. Although mostly examples were used of monological discourses, many of the principles were — often tacitly — assumed to have a more general nature. Thus, semantic and pragmatic coherence, both at the local and global level, is a property which will hold both for monologues and dialogues. Without these notions it would even be difficult to have a theoretical distinction between different, e.g. subsequent, discourses of one or several speakers.

However, it should be admitted that much of the work on discourse analysis has neglected possible *specific* properties of dialogical discourses. And indeed, it has been the ethnomethodological analysis of conversation which, first in sociology and then in linguistics, has remedied this omission. Unfortunately, there has until recently been very little integration between the two areas of discourse and conversation analysis. Our contribution to this integration in this paper will only be very modest and take place in the domain of a cognitive model.

2.6. The second area of research which constitutes the theoretical background of this paper are recent developments in cognitive psychology and artificial intelligence. In the paradigm of 'semantic information processing' attention has been extended from sentential to textual materials. Models have been designed and experiments are being carried out about the processes and representations involved in the comprehension of discourses, e.g. of stories. The contribution of artificial intelligence in this area first of all pertained to the explication of the role of knowledge in understanding, both of 'real' and of textually represented states of affairs. Notions such as 'Frame' and 'script' were developed in order to represent the organized and conventional nature of our conceptual knowledge of the world. At the same time it was shown in this approach that discourse comprehension depends on our knowledge about

---

plans and goals of (represented) participants in interaction, and about other properties of action in general. Both in cognitive psychology and in artificial intelligence it was finally recognized that discourses may exhibit various kinds of 'schematic' structures, such as the categorical organization of a story or psychological paper.

The theoretical and experimental research has been focussing mainly on processes of comprehension. Discourse production has been little studied, probably because its properties are more difficult to assess experimentally. Comprehension was mainly tested in paraphrase, free or cued recall (reproduction), recognition, question answering and summarizing tasks. It was found that especially in delayed recall subjects will no longer be able to retrieve most propositions of a discourse. They will mainly recall macrostructures (main topics) and, personally varying, salient details. Summaries appear to feature the same properties. In both cases it was assumed that during reading and comprehension language users not only understand the respective sentences of a discourse and their semantic connections, but at the same time apply various 'macrorules' which construct the global topic or gist of the text. It is this macrostructure which organizes the representation of the text in memory and which is easier and longer accessible than the 'details' of the discourse.

Similarly, conventional schemata also play an important role in the global organization of the information of a discourse, both during comprehension and storage, and in (re-)production. Experiments have shown that the presence of well-known (narrative) schemata facilitate comprehension and recall.

Both in local and global comprehension of discourse it appeared that knowledge of the world (frames, scripts) play a decisive role, e.g. in the establishment of coherence and the formation of macrostructures.

### 3. SOME SPECIFIC PROPERTIES OF DIALOGUE DISCOURSE

3.1. In order to be able to see whether the cognitive models developed in these areas can simply be accommodated to also account for processes involved in dialogue interaction, we must briefly enumerate some of the properties which distinguish monologues from dialogues. In our first section we already summarized some main levels and domains of analysis for dialogues. The question now is where the properties involved are specific, so that extensions of the cognitive model would be needed at those points.

3.2. The obvious fact that dialogues require several participants as subsequent

---

speakers has an immediate implication for experimental psychology. In current laboratory experiments textual materials are mostly used 'out of context'. That is, they are not produced, transmitted, understood and 'used' under normal pragmatic and social conditions. The pragmatic context is that of the psychological test situation, which has specific assumptions of readers/hearers, specific tasks and goals, etc. In other words, the results of experiments on discourse comprehension in this area are only an approximation of natural reading and comprehension processes. Now, in dialogues the hearer-comprehender must be at the same time a (real) participant which on a next turn becomes speaker. So, adequate experiments in this case would at least need a simulated context in which subjects engage 'naturally' in verbal interaction.

This is not only a difference and additional difficulty in experimental design but at the same time an aspect of the cognitive model itself: it needs both a comprehension *and* a production component, and there must be intimately related. Below we will assume for instance that in dialogues of certain kinds, e.g. daily conversations, processes of understanding and of planning for production may be mingled.

3.3. There are various kinds of dialogues, some of which may be close to *discourse sequences*, especially in written communication. Thus, a lecture would essentially have a monological character, but comments or questions should be considered as relevant subsequent verbal reactions to them. On the other hand, proper dialogues may consist of turns which have acquired a more or less independent nature, e.g. when we tell a story during a conversation.

Although there are several types of transitions between monologue and dialogue, we will provisionally establish dialogues as *global units of interaction*. Besides the various structural properties of the respective *utterances* constituting a dialogue (spatio-temporal continuity, grammatical connections and semantic coherence), we therefore should especially focus on the *action structure* of the interactional unit.

In the first place this means that the respective *speech acts* of the dialogue will be intended and performed on the basis of beliefs and purposes about subsequent speech acts of the hearer as the next speaker. In other words, both for speaker and hearer speech acts of a dialogue may each be planned or interpreted as a condition for the performance of speech acts in a next turn. And, similarly, each subsequent speech act will be planned and understood as a reaction to previous speech acts. So, at the pragmatic level dialogues are

---

sequences of utterances which *are connected*. This pragmatic connectedness requires the semantic and grammatical connectedness of the (semi-)sentences expressed by these utterances.

This kind of linear connectedness of speech acts performed by subsequent speakers will in several kinds of dialogues, e.g. interviews, meetings, etc., be complemented by *global pragmatic coherence*. That is, for one or more participants involved, the subsequent speech acts are at the same time performed as instantiations of one or more *global* (speech) acts. In that case the dialogue may have one over-all intended result and one over-all purposed aim or goal, being the consequence of that result, e.g. as follows: 'B has been requested by A to borrow him thousand dollars' may be the global result of a successful global request, which may lead to the global consequence 'that B will (indeed) borrow him the money'. The same may hold in the perspective of arriving at a common decision or at a confession of one of the participants. In everyday conversation such an *initio*' goal need not exist (small-talk) initially, or may have a very vague nature, such as confirmation of friendly social relationships.

We observe that dialogues are utterance sequences performed by subsequent speakers which are both locally and (often) globally coherent at the pragmatic level. Details of this pragmatic coherence, e.g. the relations between local and global coherence, will not be discussed here.

In the second place it should be noted that the local management of speech act sequences in subsequent tucos may be different from monological discourse sequences (such as storytelling rituals). That is, speakers not always have control over the successful performance of their speech acts: they may be interrupted by other participants taking a turn, which may also lead to partial overlapping of utterances. The dialogical interaction, e.g. in conversation, is characterized by a distribution of rights and obligations of speaking, whereas in a monologue the speaker has (taken) the right to control the length of the utterance, although of course there exist socially accepted conventions which limit its length in practice (a lecture may not take 6 hours, say).

Finally, what has been said for speech act sequencing and local turn management holds for the further social implications of the unit of interaction. That is, not only the linguistic and pragmatic context are permanently changing, but also the social context. First of all, each speech act and its accompanying paratextual acts (face work, gestures, etc.) will in each *stage* of the communicative social context imply a number of *further social act interpretations*. Thus, an assertion may further be intended and/or

---

interpreted as an attack, as 'showing confidence', as performing one's duty or compliance with a norm, etc. That is, each speech act or turn will change the social situation which will be the initial condition for the successful performance and interpretation of the next social act of the interaction sequence: the speaker and/or hearer have different rights, obligations, their roles may change, and the social 'Frame' may be changed. Again, these changes may take place both locally and globally. In a meeting we may have 'local' deliberations on one point, and at the same time make global decisions, whereas a trial will result in a final, globally intended social result, viz. a judgement, both having important social consequences. Whether initially planned or not, whether under the control of one or more speakers or not, this means that at the level of social analysis of dialogical interaction the dialogue can be defined in terms of identifiable differences between the initial social context and the final social context. This change is accomplished by a locally and globally coherent sequence of interactionally relevant actions of several participants.

Clearly, this is still far from an adequate definition, let alone theoretical framework, of dialogues. Our indications are not only very rough, but also were not specified for various types of dialogue, e.g. a daily conversation, a formal oral exam, a meeting, an interview, institutional 'exchanges' of various kinds (e.g. buying a ticket in the station or the movie theatre), etc. Yet, we now have some elementary insight into what might be needed in a cognitive model of dialogue.

#### 4. COGNITIVE ASPECTS OF DIALOGUE

4.1. An empirically adequate cognitive model of dialogue, as we may guess from the previous indications, will not be a simple thing. First of all, it should contain a language processor and an (inter-)action processor. It must feature much social knowledge and knowledge of the world in general. It should have both a production and a comprehension component. It must have sets of textual and social strategies, and so on. In this stage of research in cognitive psychology there is no hope to come up with such a full-fledged model in the near future. So, we must try to model and experimentally test only some manageable parts of the model.

4.2. Therefore, let us start at a point where we *do* have some insight, viz. *discourse comprehension*. That is, at any point of a dialogue a hearer needs to understand the actual utterance of the current speaker. Moreover, he must link this utterance — or rather his interpretation of it — with previous



---

utterances of the same speaker, and with his own previous utterances. That is, the internal textual structure of the current utterance must be understood, as well as the various coherence relations, both local and global, with the others of the dialogue. We will further assume that the understanding of the discourse is the decisive condition for the further understanding of the (speech or other) acts being performed by the speaker, and for the preparation of the (speech) acts of the hearer as a next speaker. Understanding of dialogue utterances, except for the first utterance, is therefore *relative*: all consequences of previous sentences and acts will count as presuppositions of interpretation.

The model, thus, first of all will contain the usual mechanisms for language processing, e.g. a parser analyzing surface structures *and* interpreting them as conceptual ('semantic') structures, e.g. propositions. Note that already at this point oral dialogue at the same time requires a mechanism for analysing and interpreting paratextual expressions. The terminal semantic interpretation may *be* the output of both analysers. We will further neglect the surface structure parsers and focus on the conceptual management of dialogues.

In a model of discourse comprehension it is assumed that, thus, sentences are 'translated' into structured proposition sequences. These are stored in *semantic working memory*. In order to connect sentences of a discourse these propositional structures must be linked, e.g. by conditional connections (cause, reason, enablement, situation-fact, etc.). This linear process of connection and coherence establishment has a *cyclical* nature: new propositions are being constructed whereas some 'old' propositions need to be stored elsewhere, e.g. in episodic memory. In this way, there takes place a gradual construction of a *representation* of the discourse in episodic memory. Since propositions as expressed by the discourse itself need not always be directly connected and coherent, world knowledge, as e.g. organized in frames or scripts in our semantic knowledge store of Long Term Memory (LTM), will supply necessary propositions to establish linear coherence.

However, the resulting representation is not sufficiently structured. During input comprehension in semantic working memory we are now able to link sentences with (previous) sentences, but not with the meaning of the discourse as a whole. That is, we do not know in which respect the sentence is relevant for the actual *topic* of the discourse. This means that *macrones* must operate at the same time and construct from input proposition sequences one or more macropropositions which represent the actual topic, and which can be kept in the buffer store of semantic working memory. At

the same time these macropropositions will further assign hierarchical structure to the discourse representation in episodic memory. The macro-rules delete irrelevant propositions, make generalizations over sequences and construct propositions denoting normal conditions, components and consequences into a more global proposition denoting a more global fact. In all these cases, again, world knowledge must be activated in order to apply the macro-rules. The resulting macrostructurally organized representation of the discourse may also be further structured by assigned *schematic* global structures, e.g. a narrative schema, or a dialogue schema.

Finally, it should at least briefly be mentioned that all these processes are controlled not only by a conventional knowledge system, but also by general values and norms, and further by the actually relevant opinions, attitudes, interests, tasks, etc. of the hearer. This collection of contextually operating and personally varying factors of processing will be called the *cognitive set* of the hearer. Note that at each point of the discourse (at each cycle) the cognitive set will change: the hearer knows more, may change opinions about the discourse or its author, form wishes, etc.

Against the background of this roughly sketched model of discourse comprehension we should see how utterances of dialogues are understood. First of all it should be recalled that such utterances will not always express complete sentences, and not even phrases or meaningful words (cf. frequent *Mhmm*'s). Yet, hearers at least roughly make sense out of them. How conceptual representations can be formed on the basis of such grammatically incomplete input is a problem we cannot deal with here. Important in this case is the powerful system of inferences, which supplies missing information from both general world knowledge and knowledge and expectations about the context (including the speaker) and what has been said before, and from the global topic and global speech act now being 'carried out'. This does not mean that each utterance will be fully interpreted and represented by the hearer: we will have to account also for more or less frequent lacks of attention, e.g. due to own speech planning or contextual phenomena, which may make interpretations, at *least locally, partial*

For longer utterances, expressing several sentences, the discourse comprehension mechanisms will operate which have been described above: macrostructurally organized representations will be constructed for each utterance and stored in episodic memory. These will be the basis for further interpretations and connections with previous and subsequent utterances.

We have seen at several points in our discussion that local comprehension is not sufficient. This means that the representations of each turn of the

dialogue should be connected to those constructed for previous ones, both of one's own utterances and of those of other participants. Connections in this case, as we have seen before, will first be linear or sequential: propositions will be linked e.g. by conditional connections, co-reference will be established between expressions, etc. We may assume that in general this relevant 'previous information' must be taken from episodic memory, because the semantic working store has too limited a capacity to contain more than a few micro- and macropropositions. Moreover, as we will see below, turns not only exhibit different sentences but also different actions, which need both comprehension and planning, which cannot simply take place with the information stored in the short term buffer. Take for instance the following simple dialogue:

- (2) A: Are you coming?  
B: OK

In order to be able to semantically link both utterances we should be aware of the fact that B should understand A's utterance, first semantically, second pragmatically, viz. as a question or request, upon which a complex decision procedure will take place, finally resulting in B's utterance. Hence, although the two turns and sentences are contiguous, there is a considerable cognitive 'distance' between them in a complete processing model. This may mean that even previous propositions must be reinstated from episodic memory to the working store.

The establishment of coherence at a more global level of comprehension would imply that the hearer determines in what respect the propositions of the current utterance belong to the same topic. This actual topic may well *be* kept in the short term store, because it is permanently needed in order to semantically *monitor* the dialogue: 'what are we talking about?'. Of course, all kinds of interruptions, side sequences or other embedded turns or discourses, topic changes and topic 'losses' may occur, which presupposes that macropropositions are no longer available in the STM store.

4.3. The nature of verbal interaction, especially in dialogues, not only requires that hearers understand what the other is 'meaning by what he is saying', but also 'what he is meaning by what he is (thereby) *doing*'. In other words, the hearer will have to assign speech act concepts to the actual utterance, that is he must also *pragmatically comprehend* the actual utterance.

The cognitive processes of pragmatic comprehension are even more

complex than those of semantic comprehension. That is, a sentence like that of A in example (2) above cannot simply be assigned the concept 'question' or 'request' by the form and the meaning of the sentence alone. Speech acts are defined in terms of various *contextual* properties, such as beliefs, wants and social relations of participants. Under specific conditions, then, (2,A) may also be a threat, an invitation, or an advice. It follows that the hearer must have or make an adequate *analysis of the relevant pragmatic context*. Speech acts, like other social actions, involve knowledge, beliefs, wishes, etc. of a speaker, and also conditions, purposes, etc. linking them to other (speech) acts. This means that the hearer has to check his representations of the previous utterances of the dialogue, of the actual context, the actual social frame, etc. in order to find the constitutive indices defining a possible pragmatic context in which some speech act may be expected and appropriate. This intricate process of comprehension will be both bottom up (given the utterance and its semantic interpretation) and top down (given previous knowledge and expectations of various kinds), and will result in a more or less firm *hypothesis* of what is meant by the speaker at this pragmatic level. Sometimes these data will be insufficient, which requires desambiguation questions: "Is that a promise or a threat?".

In the same way as sentences expressed by dialogue utterances are not comprehended in isolation, speech acts as we saw will be interpreted relative to other speech acts of the sequence. Thus, (2,B) may be interpreted by A as an adequate answer of consent. In order to be linearly connected and coherent, though, each speech act must satisfy the specific conditions which are defining the final state of the previous speech act (as in local question-answer pairs) *and* the actual global state of the global (speech) act now being performed:

(<sup>3</sup>) A: Are you coming?

B: OK. Let's go.

In the second turn of this dialogue, B not only gives a direct answer, but also initiates a speech act which is a possible next step in an interaction sequence in which the request-consent is embedded. Similarly, a hearer may produce or understand a speech act as a possible component, condition or consequence of a global speech act going on, e.g. thanks in a request ritual, or all kinds of preparations of such a request.

From this brief discussion it may have become clear that the processes involved in pragmatic comprehension are extremely complex; we know little about them and will therefore not discuss them further here.

4.4. The next component in the comprehension part of a cognitive model of dialogue is the proper *interactional* component: given the linguistic and pragmatic information of the utterance, and hence a representation of its meaning and speech act function, the hearer will at each stage of the sequence have to further establish a representation of the actual *social situation*. At least part of this analysis already takes place in order to be able to assign a pragmatic interpretation to the utterance. Some of the additional types involved however are:

- (4)
  - (i) social context type (formal/informal, public/private, etc.);
  - (ii) social (sub)system (public transport, education, etc.);
  - (iii) social frame or scripts (taking a bus, having a class, having breakfast, taking a drink in a bar);
  - (iv) participant categories (roles, functions, etc.);
  - (v) conventions of the frame or scripts (rules, norms, and their inferred expectations);
  - (vi) global social action now being performed;
  - (vii) previous local action(s) of the sequence;
  - (viii) dialogue schema categories ('opening', 'closing', etc.).

This list is not arbitrary. That is, we assume that the hearer, in order to manage the enormous complexity of data relevant for the full interpretation of the actual utterance will apply an analytic *schema*, with which the situation is analysed on crucial points for its type, (sub-)system, participant categories, etc. Again, the processes involved will be both bottom up and top down: very often the hearer already knows or expects what the relevant context and its properties are. In that case the observation and interpretation data need only be checked to the possible terminal categories of the social schema.

Note also that at this point we not only have a canonical analysis of the social situation, but also a representation of the *actual* interactional *strategy* of the speaker by which each utterance/turn is to be interpreted as a possible *optimal move* for the realization of a vague or precise goal. Again, we ignore the precise details of this process: here is another task of the model.

4.5. Having left many (also important) details aside we now have mentioned the major components in the comprehension part of the processing model. That is, we have discussed some properties of the first *stages* of that complex process of interpretation. We assumed that the various interrelated interpretations are represented in (episodic) *memory*. However, *what these*

representations look like has not been made explicit. We only assumed that the representation of the utterances is a hierarchical propositional structure, where macropropositions 'themes' or 'topics' are 'high' in the representation. The problem now is how the representation of the speech acts and other social acts, as well as other properties of the communicative situation, are represented and linked to the representation of the utterance. Is it the case, for instance, that there is a complex representation of the action(s) being performed by the speaker, e.g. 'giving an oral exam', 'asking questions', in which the representation of the (semantic) structure of the actual utterance is a functional part? Or do we have separate representations which are only systematically linked at some points? At the moment this is merely a theoretical problem for a cognitive model of comprehension, which might be decidable if experiments can be designed to establish further data for either one of these alternatives. Important though is the strong assumption that the representations *are highly structured, hierarchically organized*, at least *systematically* linked, and if possible organized according to existing cognitive *schemata* of some kind (e.g. parts of some frame or script). Without these assumptions it would, we think, not be possible for a hearer to organize the amount of information which is relevant in the interpretation of moves in dialogue sequences. Crucial in this respect is not only storage in memory but especially *retrieval*, which constitutes a necessary condition for *appropriate reactions*, i.e. for the *production of a next move*.

4.6. Although we possess little experimentally supported insight into processes of discourse and dialogue *production* — apart from some work in psycholinguistics about sentence production — we may safely assume that many of the general principles about complex information processing discussed above for comprehension also are valid for production. That is, a discourse or dialogue fragment may be uttered after morphophonological and syntactic formulation of underlying conceptual structures — taking into account the various functions of words/sentences in the text and context. Important for our discussion is that we should further assume that meanings of sentences are planned under the *monitoring global plan* of a semantic macrostructure or theme/topic. Of course this plan may be pre-programmed or not, be rather stable for part or the whole of the dialogue, or not, may be changed according to strategy, etc. The complexity of semantic coherence makes such global conceptual planning necessary. At the same time it establishes on the one hand a link with world knowledge (frames, scripts) and on the other hand with the global plans of the social actions being carried out.

We may further assume that the global planning of action is based on a complex motivational structure, involving decision making, intentions and purposes as determined by wishes, needs, knowledge and beliefs, attitudes and norms, etc. which cannot be discussed here. Given the global representation of certain (wished) goals and the representation of the (global) action to be performed to reach that goal, the action may be locally 'translated' into more specific (speech) acts, which are sequentially ordered and connected. Local data may of course change strategies and hence global plans.

Hence, whereas in comprehension global representations of meaning and action are gradually constructed in episodic memory after initial interpretation in STM, the production process first has global representations in episodic memory which by various processes (e.g. inverse macrorules and other semantic transformations) are given to working memory to be carried out. Again, the precise 'surface' details of speaking and acting will not concern us here: we first need a rough sketch of the model as a whole.

4.7. The comprehension and production processes do not simply 'follow' each other in a model of dialogue. Clearly, they are permanently interdependent. Representations resulting from comprehension are the input data for a process of action and utterance production. Important from a cognitive point of view are the striking *time limitations* in oral dialogue. A hearer cannot simply understand what is being said and what is being done, draw a number of relevant inferences, check upon his own wishes, etc., make decisions, form intentions and purposes for global plans of action or local executions thereof, etc. *after* the current utterance/doing. Answers mostly come right away, often in the middle of a current utterance. ***Hence, comprehension and production planning must at least partly occur simultaneously.*** This may be an obvious empirical fact, but the consequences for a cognitive model are not obvious at all. Short term memory is, as we observed, strictly limited in storage capacity. Although with the help of macrostructures and schemata of various kinds we are able to reduce and organize the complex information such that it can indeed be handled there, we have no insight yet in how *at the same time* details of action planning and its complex 'underlying' motivation may take place in the same working memory. This seems possible only if in each cycle of comprehension, information is not only linked to current world knowledge, but also to current other factors of the *cognitive set* (beliefs, attitudes, wants, interests, tasks, goals, etc.) and inferences drawn not only for the construction of what is being said/done by the speaker, but also what *must* be done by the hearer. Although schemata, fast strategies, macro-

planning, etc. are relevant here too, it is still striking to witness that all this takes place in milliseconds. Compared to 'simple' syntactic analysis and production studied in psycholinguistics, the processes involved here seem vastly more complex, and in the near future we will at most be able to model some small fragments of it. A cognitive theory of dialogue is a challenge to psychology and artificial intelligence precisely because both production and comprehension, both language and action, and both cognitive and social contexts are involved, all being closely interrelated, and working under specific conditions of strategic, time-limited, interaction processes. This paper, therefore, is hardly more than a sketch of some of the future tasks and problems and hardly the beginning of possible answers or solutions.

*University of Amsterdam*

#### NOTES AND REFERENCES

The remarks in this paper about the structures of dialogues and conversation, are mostly due to results from current work in sociology, anthropology and linguistics about the analysis of conversation, e.g., as done by Sacks, Schegloff, Schenkein, Turner, Sudnow, Cicourel, Franck, and many others referenced in other papers contributed to this book.

Some more specific remarks about the strategical nature of conversational interaction are due to the work of Dorothea Franck, e.g., her thesis *Grammatik und Konversation* (University of Amsterdam, 1979).

For the theoretical background in the domain of discourse theory (text grammar and other theories), cf. e.g. Wolfgang Dressler (ed.) *Current Trends in Text Linguistics* (Berlin/New York: de Gruyter, 1977) and my *Text and Context* (London/New York: Longman, 1977), and the numerous references given in these two books.

The discussion of some of the assumed properties of a cognitive model of dialogue is an extension of my work, partly in collaboration with Walter Kintsch of the University of Colorado (Boulder), on disemine processing. For the current state of our processing model and for further references to (our and other's) earlier and further work in this domain, see: Walter Kintsch & Teun A. van Dijk, 'Toward a Model of Text Comprehension and Production', *Psychological Review* 85 (1978), 363-394.

The more general theoretical background in cognitive psychology and artificial intelligence constitutes the work by scholars such as Kintsch, Rumelhart, FredenIcsen, Freedle, (Jean) Mandler, Meyer, Thorndyke, Bower, and others, on the one hand, and Minsky, Charniak, Schank and Abelson, and others, on the other hand, e.g. as appearing in the following recent books on discourse processing:

Roy O. Freedle (ed.), *Discourse Production and Comprehension* (Norwood, N.J.: Ablex, 1977).

Marcel Just & Patricia Carpenter (eds.), *Cognitive processes in Comprehension* (Hillsdale, N.J.: Erlbaum, 1977).



---

Roger Schank & Robert Abelson, *Scripts, Plans, Goals and Understanding* (Hillsdale, NJ.: Erlbaum, 1977).

For relevant further references these books may be consulted. For an interdisciplinary treatment of global structures in discourse, interaction and cognition, as mentioned in this paper, see my *Macrostructures* (Hillsdale, N.J.: Erlbaum, 1980).

*POSTSCRIPT* (MAY 1983)

This paper was written nearly five years ago. In the meantime both conversation analysis and the psychology of discourse processing have undergone rapid developments, so that much of the general and programmatic observations in this paper could now be filled in with much more detail. Especially the *strategic* nature of both conversation and cognitive understanding can now be spelled out in more precise terms, as has been done in my book with Kintsch, *Strategies of Discourse Comprehension* (New York: Academic Press, 1983), in which also further details of the cognitive model have been specified which are relevant for the cognitive modelling of dialogical interaction. In that theory discourse monitoring (by macropropositions, goals, scripts, etc.) is not handled in STM, but by a separate Control System. Also, besides the proper representation of the discourse and the interaction in episodic memory, we now also postulate a separate *situation model*, i.e. an episodic representation of the fragment of the world a discourse is about, as well as a *communicative situation model*, which is a representation of the actual interaction context of the discourse/dialogue. This CSM is fundamental for the understanding of what 'goes on', e.g. socially, during conversation. CSM features previous dialogue experiences (of the same kind), and instantiated general schemata, such as specific dialogue scripts.