These last two strategies of reference-resolution suggest that the proper direction of future research in this area should not be limited to further investigations of how people interpret pronouns in decontextualised sentence pairs, but rather should be based on more naturally occurring discourse of different types. We hope that our presentation of some of the complexities involved in the interpretation of pronominal reference will stimulate such research and discourage the reader from accepting any simplistic 'substitution' view of the function of pronouns in discourse.

7
Coherence in the interpretation of discourse

7.1 Coherence in discourse

One of the pervasive illusions which persists in the analysis of language is that we understand the meaning of a linguistic message solely on the basis of the words and structure of the sentence(s) used to convey that message. We certainly rely on the syntactic structure and lexical items used in a linguistic message to arrive at an interpretation, but it is a mistake to think that we operate only with this literal input to our understanding. We can recognise, for example, when a writer has produced a perfectly grammatical sentence from which we can derive a literal interpretation, but which we would not claim to have understood, simply because we need more information. Extract (1), the first sentence of a novel, may provide an illustration of this point.

(1) Within five minutes, or ten minutes, no more than that, three of the others had called her on the telephone to ask her if she had heard that something had happened out there.
(Tom Wolfe, *The Right Stuff*, Bantam Books, 1981)

The novelist is, of course, leading his reader to read on and find out just what the first sentence, though literally complete, has only partially described.

At the opposite extreme, we can point to linguistic messages which are not presented in sentences and consequently can't be discussed in terms of syntactic well-formedness, but which are readily interpreted. Our lives are full of such 'fragments', as in extract (2) from an Edinburgh University notice board and extracts (3) and (4) from newspaper advertisements.

(2) Epistemics Seminar: Thursday 3rd June, 2.00 p.m.
Steve Harlow (Department of Linguistics, University of York).
'Welsh and Generalised Phrase Structure Grammar'

Although it is not stated, literally, in this discourse fragment, we know that Steve Harlow (and not a person called Epistemics Seminar) will give a talk (and not write or sing or show a film) with the title shown in quotation marks; in the University of Edinburgh (no, not York, that's where he comes from); on the nearest 3rd June to the time of the notice being displayed, and so on.

(3)   Self Employed Upholsterer
        Free estimates. 332 5862.

(4)   Find the Ball. Win a House. Page 4.

If we encounter (3), we are expected to understand that the source of the advertisement is the *upholsterer* and that he or she will provide *free estimates* of the cost of upholstery work which the reader may need to be done. It is not a random assortment of words and numbers. Although it is not stated in (4), we should expect that on *page 4* of the newspaper, there will be a competition with the task determined by the first sentence and the prize detailed in the second. Despite the imperative forms, the required interpretation of the first two sentences involves the first as a condition for the second.

We might say that, in addition to our knowledge of sentential structure, we also have a knowledge of other standard formats in which information is conveyed. We also rely on some principle that, although there may be no formal linguistic links connecting contiguous linguistic strings, the fact of their contiguity leads us to interpret them as connected. We readily fill in any connections which are required.

This last point we have already mentioned in connection with the assumption of **coherence** which people bring to the interpretation of linguistic messages. Yet, the assumption of coherence will only produce one particular interpretation in which the elements of the message are seen to be connected, with or without overt linguistic connections between those elements. On the assumption of coherence, extract (3) could be interpreted as an advertisement by someone looking for an upholsterer. There is nothing in the literal message to discourage such an interpretation. There are several things in the *reader*, however, which lead him to avoid this interpretation. The most important of these is the reader's (or hearer's) effort to arrive at the writer's (or speaker's) *intended*

meaning in producing a linguistic message. We have already appealed to this notion in our discussion of discourse reference in Chapter 6. More formal arguments in support of this view of interpreting meaning can be found in Grice (1957) and Schiffer (1972).

On what does the reader base his interpretation of the writer's intended meaning? In addition to the assumption of coherence, the principles of analogy, local interpretation and general features of context, already discussed in Chapter 2, there are the regularities of discourse structure outlined in Chapters 3 and 4, and the regular features of information structure organisation detailed in Chapter 5. These are aspects of discourse which the reader can use in his interpretation of a particular discourse fragment. Yet, the reader also has more knowledge than knowledge of discourse. He knows for example, that *Steve Harlow* is much more likely to be the name of a person than *Epistemics Seminar*. This is a form of conventional socio-cultural knowledge. He also knows that the purpose of the linguistic message, its function in communicative terms, is that of an announcement and not a warning (or a promise, or whatever) partly because of its location, partly because of its form, and partly because of the same socio-cultural knowledge that leads him to know what are, and what are not, usual names for people. He may, of course, have some highly specific local knowledge, deriving from the fact that he is a linguist, has met Steve Harlow, knows of his interest in Phrase Structure Grammar and so on. On the basis of this, he may infer that Steve Harlow is going to use the Welsh language to exemplify certain aspects of Phrase Structure Grammars in the way that Gazdar (another element of specific local knowledge) has done with English. With such an inference, the reader may be said to have gone beyond the discourse-producer's intended message. As we shall demonstrate, however, there is a wide range of possible inferences made by readers in interpreting discourse and it is not always easy to determine which were intended by the text-producer and which were not.

We have isolated three aspects of the process of interpreting a speaker's / writer's intended meaning in producing discourse. These involve computing the communicative function (how to *take* the message), using general socio-cultural knowledge (facts about the world) and determining the inferences to be made. We shall

discuss these in more detail in the course of this chapter and consider proposals which have been made to account for these aspects of discourse understanding.

7.2 Computing communicative function

As we pointed out in Chapter 1, there has been a long tradition among social anthropologists and ethnographers studying the language of speech communities of assuming that speakers convey both social and propositional meanings when they produce particular utterance forms in particular contexts. (For relatively early work in this tradition, see Malinowski, 1935.) In recent years there has been a development of interest in the 'social interaction' aspect of language use. Much of this work has been carried out by sociolinguists who have attempted to describe how an utterance can 'count as' a social action such as a greeting or a promise or, in the case of extract (2), an announcement rather than a warning. We shall consider some brief examples used to support the view that utterances must be treated as 'actions' of different types and review the theoretical and descriptive frameworks developed in support of this approach.

Labov (1970) argues that there are 'rules of interpretation which relate what is said to what is done' and it is on the basis of such social, but not linguistic, rules that we interpret some conversational sequences as coherent and others as non-coherent. As an example of a non-coherent conversational sequence, Labov quotes the following example of a doctor talking to a schizophrenic patient, from Laffal (1965: 85).

(5)   A: What's your name?
        B: Well, let's say you might have thought you had something
           from before, but you haven't got it any more.
        A: I'm going to call you Dean.

Labov points out that the recognition of coherence or incoherence in conversational sequences is not based on a relationship between utterances, but 'between the actions performed with those utterances'. Other analysts have attempted to develop this point, frequently basing their discussions on examples such as (6) and (7).

(6)   A: What time is it?
        B: Well, the postman's been already.

This example, quoted in Brown & Levinson (1978: 63), is used to show that the assumption of rationality on B's part leads us to assume that he is providing an answer to the question asked, and so on to the conclusion that the time is past 11 a.m., for example. The next example is taken from Widdowson (1979a: 96) and illustrates a coherent piece of conversational discourse which exhibits no cohesive links between the two sentences involved.

(7)   A: Can you go to Edinburgh tomorrow?
        B: B.E.A. pilots are on strike.

Widdowson claims that B's reply is to be taken as a negative answer to the question, because the strike will prevent the speaker flying to Edinburgh. This is clearly one interpretation of the speaker's intended meaning but we could also suggest others; for example, that the speaker intends a 'don't know' response because he is not yet sure whether he will try some alternative transport. Whatever the intended meaning, we are in no doubt that B's utterance *counts as* a response and not just a gratuitous statement about the way the world is.

The use of some linguistic elements, such as the conjunction *because* ('*cause*) in the following two extracts, would be claimed to be explainable only in terms of an utterance-as-action analysis.

(8)   A: but you'd have telephones around
        B: mm – oh yes . . . I've had the telephone since nineteen
           thirty eight (hmm) oh they were on a long while I think
           before that
        A: 'cause there was a man in . . .

(9)   What's the time, because I've got to go out at eight?

Example (8) is taken from a recorded conversation which was presented more fully as (12) in Chapter 3. Example (9) is quoted in Levinson (1980: 8). The second example was used by Levinson to demonstrate that a conjunction like *because* is not only used to connect two clauses in a complex sentence. It can also be used to introduce the reason for asking a question, as in (9), or for introducing a particular subject into a conversation, as in (8). In other words, the structure of the above examples is not that normally associated with *because* as a logical connector (P because Q), but is as follows:

I mention / ask   P   because   Q.

Consequently, our understanding of examples (8) and (9) is based, not on an interpretation of the sentences-on-the-page, but on our assumption that a reason is being expressed for an *action* performed in speaking. The action, and the reason for it, are to be identified by virtue of their location within a conventional structure of spoken interaction. This conventional structure provides an account of how some utterances which are apparently unconnected in formal terms (lack cohesion) may be interpreted within a particular genre of spoken interaction, say conversation, as forming a coherent sequence. Widdowson (1978: 29) presents the following example:

(10)   A: That's the telephone.
         B: I'm in the bath.
         A: O.K.

Widdowson suggests that it is only by recognising the action performed by each of these utterances within the conventional sequencing of such actions that we can accept this sequence as coherent discourse. The conventional sequencing may be presented as in (11):

(11)   A requests B to perform action
         B states reason why he cannot comply with request
         A undertakes to perform action

Such a representation yields a description of conversational discourse as a form of social interaction. A similar analysis could be applied to a series of gestures, as in (12):

(12)   (domestic evening scene: husband and wife watching television)
         A indicates by pointing and tapping his ear that he can hear the telephone
         B points to the cat asleep on her lap
         A shrugs and gets up

The analysis of the *interaction* can be made without taking account of the language employed by the speakers. It is typical of many discussions of discourse structure which rely on an analysis of sequences of actions, that rather little attention is paid to the linguistic aspects of the realisations of these actions. In discussing discourse structure in these terms, Coulthard (1977: 7) argues that 'the structure, or constraints on the next speaker, cannot be

expressed in *grammatical* terms . . . the linguistic *form* of the utterance is almost irrelevant' (original emphases). A rather similar view is taken by Sinclair & Coulthard (1975: 13): 'the level of language function in which we are centrally interested is . . . the level of the function of a particular utterance, in a particular social situation and at a particular place in a sequence, as a specific contribution to a developing discourse'. Sinclair & Coulthard are concerned to examine the structure of discourse in classroom interaction. They identify five discoursal categories: lesson, transaction, exchange, move, act. Whereas it is, in principle, possible that they could identify some forms of utterance which characterise the boundaries of a lesson (59–60), it is clear that no forms which are unique to 'lessons' exist. 'Lesson' is clearly a sociologically determined category rather than a linguistically determined category. The 'discoursal' category 'transaction' is described thus (1975: 25):

(13)   a. there must be a preliminary move in each transaction
         b. there must be one medial move, but there may be any number of them
         c. there can be a terminal move but not necessarily.

It is clear that the structure represented here would cover a 'transaction' like that represented in (11) which, as we showed in (12) may equally well be used to discuss the structure of non-linguistic social interaction. The categories 'exchange, move and act' can be demonstrated to apply satisfactorily to an analysis of a non-linguistic interaction like a tennis match. They could be used to describe 'the winning of a point' / 'the service' or 'the volley' / and 'the act of serving' or 'the act of hitting a backhand return', respectively. This breadth of application may yield categories which are useful in the investigation of structure in social behaviour, and certainly illuminates the incidence of distributions of propositionally contentless items like *well* and *now*. It is not clear, however, that this creation of a complex taxonomy serves to *illuminate our understanding of how participants in an interaction understand what the speaker means by what he says as well as a general appeal to Grice's maxims and the principles of analogy and local interpretation would do.*

A more promising approach to the problem of social meaning

from the discourse analyst's point of view, is offered by a consideration of that area of conversation analysis which investigates **turn-taking**. The most influential work in this area is reported in Sacks et al. (1974), Schegloff (1968), Schegloff & Sacks (1973), Jefferson (1972, 1973), and, more recently, in Schenkein (ed.) (1978). The aim of this type of analysis of conversational discourse is to <u>identify the regularities of conversational structure by describing the ways in which participants take turns at speaking</u>. There are some easily identifiable regularities in the ordering of those two-turn units described as **adjacency pairs**. These can take the form of Greeting–Greeting, as in (14) or Question–Answer as in (15).

(14)   A: Hello.   (15)   A: How are you?
        B: Hi.   B: Fine.

With this type of data, the notion of the 'turn' as a unit of analysis seems quite reasonable. However, most conversational data consists of more substantial 'turns' in which several utterances can occur, or in which the basic adjacency pair organisation is difficult to determine. In extract (16), we might suggest that some of the interrogative forms function as both answers and questions, and that the final declarative form is not, in fact, an answer to any of the questions.

(16)   George: Did you want an ice lolly or not?
        Zee: What kind have they got?
        George: How about orange?
        Zee: Don't they have Bazookas?
        George: Well here's twenty pence + you ask him

The structure of this extract could be partially characterised as a *sequence*, following a suggestion by Schegloff (1972) that the adjacency pair structure can be disrupted by an 'insertion sequence' which delays the answer-part to one question-part of a pair until another answer to a different question has been provided. This is intuitively reasonable, but the immediate question which springs to mind is how does the analyst determine when an interrogative form counts as a question in an adjacency pair, or as a part of an insertion sequence, or even, as an answer? This type of question is never really raised by those undertaking the analysis of **conversational interaction**, largely because little attempt is made to discuss the relationship between linguistic form and the interactive functions

proposed. As Coulthard (1977: 92) points out, the work of Sacks, Schegloff & Jefferson produces many interesting insights into the workings of conversation, but the analytic methodology and categories employed remain so informal and imprecise that they are difficult for others to use in any practical way. The most that the discourse analyst might gain from the conversational interaction approach to an example such as (7), quoted earlier in this chapter, is that its coherence partially depends on our expectation that, according to the adjacency pair formula, what follows a question should be treated as an answer to that question. This may seem a rather obvious point to make, but it is exactly the type of point that, because of its obviousness, is rarely made explicit in the analysis of language. It captures one important aspect of how we assume that two formally unconnected utterances placed together form a coherent piece of discourse. They do so because there is an assumed coherent structure to discourse over and above the more frequently described structure of sentential form.

7.3 Speech acts

In 7.2 we discussed approaches to the identification of social meaning in terms of the activity performed by a speaker in uttering, with respect to analyses which identify actions in terms of the conventionally structured sequences in which they occur. In this section we turn to discuss the notion of **speech act** which has developed from the work of linguistic philosophers.

Speech act theory originates in Austin's (1962) observation that while sentences can often be used to report states of affairs, the utterance of some sentences, such as (17) and (18) must, in specified circumstances, be treated as the performance of an act:

(17)   I bet you sixpence it will rain tomorrow.

(18)   I name this ship the *Queen Elizabeth*.

Such utterances Austin described as 'performatives' and the specified circumstances required for their success he outlined as a set of 'felicity conditions'. More precisely, utterances such as (17) and (18) are examples of *explicit* performatives which are not just a specialised group of ritual sentence forms, but are a subset of the utterances in the language which can be used to perform acts.

Another subset are utterances which can be described as implicit performatives, as in examples (19) – (22):

(19)   Out!

(20)   Sixpence.

(21)   I'll be there at 5 o'clock.

(22)   Trespassers will be prosecuted

None of these examples contains a performative verb, but (19) can be used by a cricket umpire to perform an act of dismissal, (20) by a card-player to make a bet, (21) by anyone to make a promise and (22) by a landowner to issue a warning. By extension, it became possible to suggest that in uttering any sentence, a speaker could be seen to have performed some act, or, to be precise, an **illocution-ary act**. Conventionally associated with each illocutionary act is the *force* of the utterance which can be expressed as a performative such as 'promise' or 'warn'. Austin also pointed out that, in uttering a sentence, a speaker also performs a **perlocutionary act** which can be described in terms of the effect which the illocutionary act, on the particular occasion of use, has on the hearer.

This is an extremely brief summary of the basic elements in what has been developed since Austin, by Searle (1969, 1979) and many others, as Speech Act theory. Searle (1975) also introduces a distinction between direct and indirect speech acts which depends on a recognition of the intended perlocutionary effect of an utterance on a particular occasion. Indirect speech acts are 'cases in which one illocutionary act is performed indirectly by way of performing another' (1975: 60). Thus, example (23) can be seen as, at one level, a question about the hearer's ability, but, at another level, a request for action.

(23)   Can you speak a little louder?

A sentence such as (23), though interrogative in form, is conven-tionally used, as Searle points out, to make a request. For a recent survey of the outstanding issues in Speech Act theory, see Levinson (1980, forthcoming).

The principle interest of Speech Act theory, for the discourse analyst is, as we suggested in 7.2, that it provides an account of how

some apparently formally unconnected utterances go together in conversational discourse to form a coherent sequence. There are, however, a number of general problems with the application of Speech Act theory in the analysis of conversational discourse. An important practical drawback is expressed by Levinson (1980: 20) in the following terms: 'If one looks even cursorily at a transcribed record of a conversation, it becomes immediately clear that we do not know how to assign speech acts in a non-arbitrary way.' The problem with identifying speech acts should not necessarily lead the analyst to abandon their investigation. Rather, it should lead the analyst to recognise that the way speech acts are conventionally classified into discrete act-types such as 'request', 'promise', 'warn', etc. may lead to an inappropriate view of what speakers do with utterances. From the speaker's point of view several sentences (or syntactic chunks) strung together may constitute a single act. Thus, a fairly extended utterance may be interpreted as a warning or as an apology. On the other hand, one utterance may perform several simultaneous acts. Consider the following utterance of a husband to his wife:

(24)   Hey, Michele, you've passed the exam.

He may be 'doing' several things at once. He may be simultaneously 'asserting', 'congratulating', 'apologising' (for his doubts), etc. As it is presently formulated, Speech Act theory does not offer the discourse analyst a way of determining *how* a particular set of linguistic elements, uttered in a particular conversational context, comes to receive a particular interpreted meaning.

7.4   **Using knowledge of the world**

We might say that the knowledge we possess as users of a language concerning social interaction via language is just one part of our general socio-cultural knowledge. This general knowledge about the world underpins our interpretation not only of discourse, but of virtually every aspect of our experience. As de Beaugrande (1980: 30) notes, 'the question of how people know what is going on in a text is a special case of the question of how people know what is going on in the world at all'.

We suggested, in Chapter 2, that the interpretation of discourse is based to a large extent on a simple principle of analogy with what we have experienced in the past. As adults, we are liable to possess

quite substantial amounts of background experience and knowledge. How do we organise all this knowledge and activate only limited amounts when needed? We shall consider proposed answers to this question in section 7.6. Before we investigate this area, we shall try to clarify how this view of discourse-understanding via the use of 'world-knowledge' stands in relation to the view of literal interpretation via the 'words-on-the-page'.

7.5 Top-down and bottom-up processing

One metaphor for the way we process and comprehend discourse comes from computational modelling of language understanding. We can think of our processing of incoming discourse as the combination of (at least) two activities. In one part of the processing, we work out the meanings of the words and structure of a sentence and build up a composite meaning for the sentence (i.e. **bottom-up processing**). At the same time, we are predicting, on the basis of the context plus the composite meaning of the sentences already processed, what the next sentence is most likely to mean (i.e. **top-down processing**).

Since the main thrust of analysis in general linguistics has been towards developing a grammatical description of sentence form and meaning, any view taken on the processing of sentences has tended to be primarily of the 'bottom-up' type. A similar view can be found in some **Artificial Intelligence** (AI) approaches to linguistic data, in which the aim of the research is to develop a *parser* to analyse acceptable English sentences. In both these approaches, a sentence containing a grammatical error is rejected, rather than given a plausible interpretation. If a machine with a fully operational sentence grammar of English was presented with the following text (25) to parse, it would tend to come to a halt very quickly and return a 'non-grammatical' or 'unacceptable' reading for the sentence in the second line. The grammarian or AI researcher who designed the machine's program would be pleased with this result because, after all, the machine has fulfilled its designated function admirably.

(25)   *Slim is beautiful*
         Many reasons are there for peaple to want a slim body. All become very lighter and lighter but it's very diffecult to held a normally weight.

Nowadays, in our country, Sweden, there is so well of all sort of eating that man light come to big overweight. What to doing?

(We are grateful to Gunnel Melchers of the English Department, University of Stockholm, who brought this text to our attention.)

However, human processors, unlike the machine parser, do not reject ungrammatical text, they try to interpret it. We suspect that the reader has a reasonable interpretation for the writer's intended message in the discourse fragment (25). What enables the human processor to do this? A partial answer to this question is that the human processor does indeed 'parse' the sentences of the encountered text. It would be absurd to suggest that when we read the first line of (25) we do not attempt to build (i.e. from the bottom-up) some composite meaning for the three-word string on the basis of its structure and the meaning of the lexical items involved. At the same time, however, we suggest that the reader is also operating a top-down interpretive strategy which creates expectations about what is likely to come next in the text. (In Chapter 4 we demonstrated how effectively titles provide an interpretive point of departure for texts.) It is the predictive power of top-down processing that enables the human reader to encounter, via his bottom-up processing, ungrammatical or mis-spelt elements in the text and to determine what was the most likely intended message.

An immediate question arises. If we believe that bottom-up processing operates with rules of the sort presented in descriptions of sentential syntax and lexical semantics, what is the basis of top-down processing? A part of the answer to this question was presented in Chapter 2 where we suggested that discourse context creates expectations relating to discourse content. Another part of the answer is that once we start processing a discourse fragment we do not treat it as the first piece of discourse we have ever encountered. We have our experience of having processed other, perhaps very similarly titled, discourse fragments before. We can also draw on our experience of the way the world is – our background knowledge. Yet, as has been noted already, we amass colossal amounts of 'knowledge' and 'experience' in our lives. If top-down processing depends on our activating only a small part of this background knowledge at a time, then there must be some way in which that knowledge is organised and stored to allow easy

access. Attempting to represent the way in which background knowledge is held in mental storage has been the goal of a substantial amount of research in recent years.

7.6 Representing background knowledge

There have been several attempts to provide conventional or stereotypic representations of 'knowledge of the world' as a basis for the interpretation of discourse. These representations, found in psychological and computational approaches to discourse understanding, are mainly used to account for the type of predictable information a writer / speaker can assume his hearer / listener has available whenever a particular situation is described. Given one particular situation, such as a restaurant scene, the writer / speaker should not have to *inform* his reader / hearer that there are tables and chairs in the restaurant, or that one orders and pays for the food consumed therein. Knowledge of this sort about restaurants is generally assumed. In representations of this knowledge, conventional aspects of a situation, such as the tables and chairs in a restaurant, can be treated as **default** elements. These default elements will be assumed to be present, even when not mentioned, unless the reader / hearer is specifically told otherwise. A good example of our ability as readers to provide default elements automatically was demonstrated in the consideration of the recipe text, (22) in Chapter 5.

It is a feature of these knowledge representations that they are organised in a fixed way as a complete unit of stereotypic knowledge in memory. Thus, knowledge of a restaurant scene is treated as being stored in memory as a single, easily accessible unit, rather than as a scattered collection of individual facts which have to be assembled from different parts of memory each time a restaurant scene is mentioned. This aspect of knowledge representation is generally in line with a related characteristic of the approaches we will describe, insofar as they all treat discourse understanding as a processing of information in memory. Riesbeck (1975), for example, boldly asserts that 'comprehension is a memory process'. Understanding discourse is, in this sense, essentially a process of retrieving stored information from memory and relating it to the encountered discourse. An important direction of the research in this area has consequently been towards finding the best storage

concept for handling the pre-existing conventional knowledge. It should be noted that with this emphasis on the form of 'storage', little attempt has been made to demonstrate how the information stored in memory is *learned*. If it should turn out to be the case that the way we use stored knowledge is in some way determined by how we come to have that knowledge, then it is possible that the concept of a fixed storage system will have been rather misleading.

The emphasis on storage of knowledge-of-the-world is most apparent in computational approaches to discourse understanding. In order to provide a computer with the background knowledge required to 'understand' discourse, many workers in Artificial Intelligence attempted to create large, fixed data-structures, or memories, in which knowledge was organised and stored. It quickly became apparent that generalised knowledge about the world was too large and too diffuse to be incorporated, in any encyclopaedic fashion, within the computer's memory. The answer, for some AI investigators, was to produce specialised knowledge structures for coping with discourse requiring a particular type of knowledge. That is, knowledge-of-the-world could be incorporated if the 'world' was an extremely limited one. A 'world' consisting of a fixed number of coloured blocks and other shapes is one example (see Winograd, 1972), and that of a travel agent called GUS, arranging flights in California, is another (see Bobrow et al., 1977). It then became possible to think of knowledge-of-the-world as organised into separate but interlinked sets of knowledge areas which, taken together, would add up to the generalised knowledge that humans, in comprehending discourse, appear to use. This is intuitively a very reasonable idea since, when we read a piece of text, we presumably only use that limited subset of our knowledge which is required for the understanding of that text. In other words, when we read a story involving a visit to the dentist, we use our knowledge of dentist-visiting, but not normally our knowledge of typing a letter or going to a birthday party — that is, unless some part of the text also requires that other particular subset of our knowledge to be involved.

We shall consider two AI proposals for dealing with the organisation of knowledge in memory, those relating to **frames** and **scripts.** We have selected these two because they have been very influential in considerations of how discourse is understood, and

because they are generally representative of a very large body of research in this area. (For more general discussions of this research, see Wilks, 1977; Winston, 1977; Findler (ed.), 1979; Metzing (ed.), 1979.)

We shall also consider some related attempts in psychological research to provide ways of representing knowledge stored in memory and how it relates to discourse processing. The emphasis in this area is typically less storage-oriented and more concerned with how background knowledge is used in on-line processing. We will briefly discuss **scenarios** and the much more widely used term **schemata**. The idea of **mental models** is also discussed. Although there appear to be many different terms employed by different researchers, there is a very large area of overlap in what these different terms are used to describe (see Tannen, 1979). It should be recognised that, generally, the use of different terminology and considerations of different types of knowledge in these various research areas do not represent sets of competing theories. The different terms are best considered as alternative metaphors for the description of how knowledge of the world is organised in human memory, and also how it is activated in the process of discourse understanding.

7.6.1 *Frames*

One way of representing the background knowledge which is used in the production and understanding of discourse can be found in Minsky's *frame-theory*. Minsky proposes that our knowledge is stored in memory in the form of data structures, which he calls 'frames', and which represent stereotyped situations. They are used in the following way:

> When one encounters a new situation (or makes a substantial change in one's view of the present problem) one selects from memory a structure called a *Frame*. This is a remembered framework to be adapted to fit reality by changing details as necessary.
>
> (Minsky, 1975)

It should be noted that Minsky's discussion is not primarily an investigation of linguistic phenomena (much of it is concerned with visual perception and visual memory) but is directed towards a way of representing knowledge. Since one kind of knowledge is knowledge of a language, then there are frames for linguistic 'facts'. For

example, Minsky draws an analogy between a frame for a room in a visual scene and a frame for a noun phrase in a discourse. Both frames have obligatory elements (wall / nominal or pronominal) and optional elements (decorations on the walls / a numerical determiner). The basic structure of a frame contains labelled *slots* which can be filled with expressions, *fillers* (which may also be other frames). For example, in a frame representing a typical HOUSE, there will be slots labelled 'kitchen', 'bathroom', 'address', and so on. A particular house existing in the world, or mentioned in a text, can be treated as an *instance* of the house frame, and can be represented by filling the slots with the particular features of that individual house. Formulated in this way, a frame is characteristically a fixed representation of knowledge about the world. Some AI researchers state this point explicitly: 'I take a frame to be a static data structure about one stereotyped topic' (Charniak, 1975: 42). Others view the frame as a computational device which not only stores data, but is capable of implementing programs, that is, 'for organising the processes of retrieval and inference which manipulate the stored representations' (Hayes, 1979).

At a very general level, the notion of a 'frame' provides an attractive metaphor for thinking about discourse understanding as, at least partially, 'a process of fitting what one is told into the framework established by what one already knows' (Charniak, 1979). Thus, if you receive a postcard telling you where you should go to register your vote in a local government election, your 'understanding' of this received information can be described in terms of a 'voting-frame', perhaps, which has a slot for 'voting-place'. The specific locational information (*St Bernard's Centre*) on the card instantiates the stereotypic locational information slot in your knowledge frame. Similarly, when you look at the rest of the discourse on this postcard you see further evidence of information pertaining to your 'voting-frame', as in (26).

(26)   When you go the polling station tell the clerk your name and address.
         (Lothian Regional Council Election Poll Card, May 1982)

The definite noun phrases derive from the same 'voting-frame', in that your stereotypic knowledge of voting provides for a place to vote (*the polling station*) and an official (*the clerk*) in that place. In

other words you do not have to be informed that there is such a thing as a *polling station* and that a *clerk* will be there. The producer of this piece of discourse expects you to have this knowledge, and Minsky's frame-theory provides an account of how this expectation influences the discourse produced.

There is, however, a problem with this rather neat account of how the piece of discourse in (26) is understood. If it is indeed the case that the producer of this discourse expected the reader to process it on the basis of a stereotypic voting-frame, then one might ask why he produced the discourse at all. If you do not have to be informed of the existence of the polling station and the clerk, because you have stereotypic knowledge of these things, then why do you have to be informed of the actions you should perform? Surely your voting-frame has stereotypic actions as well as stereotypic entities. If that is the case, then you need not be given the information in (26) at all. It is an unfortunate, but nevertheless logical outcome of a frame-theory version of how we use our stored knowledge, that it predicts that a lot less human discourse should occur than actually occurs. There are many situations in which discourse is produced where the intended audience can be expected, but not guaranteed, to have stereotypic knowledge of what is to be communicated. Discourse producers, like the writer of (26), make their discourse reflect this fact, and present the information in a form which serves as a reminder for those who already know and as an instruction for those who do not.

A second, unresolved problem for what Wilks (1979) describes as 'frame-using systems', concerns the fact that, when an understander system uses a text cue to activate a frame, there may be several frames activated. Remember Minsky's proposal that 'when one encounters a new situation, one selects from memory a structure called a frame'. Consider the following new situation which presented itself at the beginning of a newspaper article.

(27)   The Cathedral congregation had watched on television monitors as Pope and Archbishop met, in front of a British Caledonian helicopter, on the dewy grass of a Canterbury recreation ground.

(*The Sunday Times*, 30 May 1982)

The problem should be immediately obvious. Is a 'Cathedral' frame

selected? How about a 'television-watching' frame, a 'meeting' frame, a 'helicopter' frame, a 'recreation-ground' frame? These questions are not trivial. After all, it probably *is* necessary to activate something like a 'recreation-ground' frame in order to account for the definite description *the grass* mentioned in the text. Yet a substantial part of such a frame, possibly incorporating a large number of sub-frames covering endless aspects of our stereotypic knowledge of 'recreation', would have no function in our understanding of this piece of text. As Wilks (1979: 153) says, 'many frames are called, but few can be chosen'.

Despite these problems, and criticisms that frame-theory is 'little more than a cumbersome convention for the listing of facts' (Dresher & Hornstein, 1976: 357), the basic concept of frames as structured repositories for our conventional knowledge has provided a useful working model for analysts, not only in AI, but also in sociology (e.g. Goffman, 1974) and linguistics (e.g. Fillmore, 1975; Gensler, 1977).

7.6.2   *Scripts*

The notion of a script was developed by analogy with Minsky's frame, but 'specialised to deal with event sequences' (Schank & Abelson, 1977). The script concept was used by Abelson (1976) to investigate the relationship between attitudes and behaviour but, when applied to text understanding, it incorporates a particular analysis of language understanding proposed by Schank (1972) as **conceptual dependency**.

Schank set out to represent the meanings of sentences in conceptual terms by providing, for any sentence, a conceptual dependency network called a *C-diagram*. A C-diagram contains concepts which enter into relations described as dependencies. There is a very elaborate, but manageable, system of semantic primitives for concepts, and labelled arrows for dependencies which we shall not describe here (see Schank 1972, 1973, for detailed discussion). We shall simply consider one of Schank's sentences and his non-diagrammatic version of the conceptualisation underlying that sentence. Examples (28) and (28a) are taken from Schank (1973).

(28)   John ate the ice cream with a spoon.

(28a)   John ingested the ice cream by transing the ice cream on a spoon to his mouth.

> The term 'transing' is used here to mean 'physically transferring'. See Schank (1973) for a fuller discussion.

One benefit of Schank's approach should be immediately clear. In his 'conceptual' version (28a) of the sentence (28), he has represented a part of our understanding of the sentence which is not explicit in the sentence-on-the-page, that the action described in (28) was made possible by 'getting the ice cream and his mouth in contact' (1973: 201). In this way, Schank incorporates an aspect of our knowledge of the world in his conceptual version of our understanding of sentence (28) which would not be possible if his analysis operated with only the syntactic and lexical elements in the sentence.

In a development of the conceptual analysis of sentences, Riesbeck & Schank (1978) describe how our understanding of what we read or hear is very much 'expectation-based'. That is, when we read example (29), we have very strong expectations about what, conceptually, will be in the x-position.

(29)   John's car crashed into a guard-rail.
        When the ambulance came, it took John to the x.

Riesbeck & Schank (1978: 252) point out that our expectations are conceptual rather than lexical and that different lexical realisations in the x-position (e.g. *hospital, doctor, medical centre*, etc.) will all fit our expectations. Evidence that people are 'expectation-based parsers' of texts is provided by the fact that we can make mistakes in our predictions of what will come next. The example (9) from Chapter 2: *John was on his way to school*, which first suggested John was a schoolboy, then later, that he was a teacher, is a good illustration of this point. Riesbeck & Schank provide the following example:

(30)   a. We went on a hunting expedition.
        b. We shot two bucks.

In our conceptualisation of this 'text', we no doubt have rifles and bullets and dead animals. We would expect the text to continue in this vein. But when we come to the third sentence (30c), we find

that our predictions were wrong and have to go back and re-fashion our conceptualisation,

c. That was all the money we had.

In analysing stories, Riesbeck & Schank supplement the conceptual analysis of sentences with a more general understanding device described as a *script*, which has a function similar to a Minskyan frame. Whereas a frame is generally treated as an essentially stable set of facts about the world, a script is more programmatic in that it incorporates 'a standard sequence of events that describes a situation' (1978: 254). (For a detailed discussion, see Schank & Abelson, 1977.) One application of a script is in the 'understanding' of newspaper stories about car accidents. Evidence of a computer's 'understanding' of such stories through the application of the script procedure is presented in the capacity to answer questions about a story. Given the story in (31), the computer can answer the questions which follow. Note that the answer to question 1 requires the machine to decide that *the passenger* and *David Hall* are the same individual and that the answer to question 2 is the result of an inference that if a person is treated and released from hospital, then he is *hurt* or *slightly injured*.

(31)   Friday evening a car swerved off Route 69. The vehicle struck a tree. The passenger, a New Jersey man, was killed. David Hall, 27, was pronounced dead at the scene by Dr Dana Blanchard, medical examiner. Frank Miller, 32, of 593 Foxon Rd, the driver, was taken to Milford Hospital by Flanagan Ambulance. He was treated and released . . .

        Q1: Was anyone killed?
        A1: YES, DAVID HALL DIED.
        Q2: Was anyone hurt?
        A2: YES, FRANK MILLER WAS SLIGHTLY INJURED.

These answers may seem trivially successful to a human understander, yet they would not normally be a product of any analysis which operated on only the syntax and lexis of the sentences in the text. In very simple terms, it is not stated in the text that Frank Miller was hurt, so how does the computer (or any other processor) come to know this? It uses a limited subset of its knowledge of the world applied to the piece of text it encounters. Riesbeck & Schank

suggest that we do the same, and that their expectation-based analysis presents 'a viable theory of how humans process natural language' (1978: 290).

Criticism of the claims of Schank and his co-authors could be made in similar terms to those against Minsky, noted earlier. That is, if scripts are stereotypic event-sequences, then would a stereotypic car crash be described at all, since we already have the information in our scripts? The problem of idiosyncratic scripts – e.g. Schank's daughter asking if he was going to get a new key chain to go with his new car (Schank & Abelson, 1977: 68) – is touched on, but not considered at length. It may be, of course, that we all have more idiosyncratic scripts than stereotypic ones.

One very specific and serious criticism of Schank's conceptual-dependency theory has been made by Dresher & Hornstein, (1976). Schank states the following condition on the well-formed-ness of conceptualisations:

A C-diagram that contains only the sententially realised information will not be well-formed conceptually. That is, a conceptualisation is not complete until all the conceptual cases required by the act have been explicated.

(1972: 569)

Dresher & Hornstein quite justifiably point out that such a condition is a recipe for endless conceptualisations. If we bring *John's mouth* into the conceptualisation of sentence (28), quoted earlier in this chapter, do we not also bring in *John's hand, his fingers, his arm muscles, his thought processes*, and so on, to arrive at a *complete* conceptualisation? This is a serious criticism and raises a problem which exists for virtually every attempt to incorporate world-knowledge in the understanding of discourse. We can see how *some* extra-linguistic knowledge is involved in our understanding, or our conceptualisation, of sentences and we can propose ways of incorporating that knowledge in our analysis. What we have difficulty with is restricting that knowledge to only the relevant details required in the understanding of particular sentences on particular occasions. The outstanding problem for Schank's theory (and for Minsky, too, as we noted earlier) is to find a *principled* means of limiting the number of conceptualisations required for the understanding of a sentence. In more general

terms, we require a principled way of constraining the expansion of any analysis which incorporates extra-linguistic knowledge in its account of the understanding of linguistic data.

Despite this general criticism of the theoretical principles involved in using 'scripts', some empirical research has shown that treating scripts as 'action stereotypes' (Bower et al., 1979) for people's knowledge of routine activities can produce experimental results to support the views of Schank and his collaborators. Bower et al. (1979) found that when they asked subjects to recall texts involving routine activities (e.g. Going to a Restaurant, Grocery Shopping, Visiting a Doctor), their subjects tended to confuse in memory actions that were stated in the text with actions implied by the 'script'. They also found that, when presented with scrambled texts which caused script-actions to be out of predictable sequence, subjects recalled the texts with script-actions in their canonical order. There is, then, some evidence that the script-concept may have some psychological validity, over and above its function as an organisational device in computer data storage. Further evidence is provided by Sanford & Garrod (1981) who base their notion of *scenario* very much on Schank's script concept.

7.6.3 *Scenarios*

Sanford & Garrod (1981) choose the term *scenario* to describe the 'extended domain of reference' which is used in interpreting written texts, 'since one can think of knowledge of settings and situations as constituting the interpretative scenario behind a text'. Their aim is to 'establish the validity of the scenario account as a psychological theory' (1981: 110) in opposition to the proposition-based theory of Kintsch (1974) which we described earlier in Chapter 3. According to the proposition-based approach, the existence of *a waiter*, for example, in the mental representation which a reader has after reading a text about *Going to a Restaurant*, depends entirely on whether a waiter was explicitly mentioned in the text. According to the scenario account, a text about *Going to a Restaurant* automatically brings *a waiter* slot into the representation. As evidence that certain 'role' slots are activated in scenarios, Sanford & Garrod show that substantial differences are recorded in the reading times for the target sentences in the following two conditions:

(32)   a. *Title: In court*
Fred was being questioned.
He had been accused of murder.

Target: The lawyer was trying to prove his innocence.

b. *Title: Telling a lie*
Fred was being questioned.
He couldn't tell the truth.

Target: The lawyer was trying to prove his innocence.

In condition *a*, with the *In court* scenario activated, reading times for the target sentence containing *The lawyer* were substantially faster than in the *b* condition where a non-specific scenario had been activated.

Sanford & Garrod emphasise that the success of scenario-based comprehension is dependent on the text-producer's effectiveness in activating appropriate scenarios. They point out that 'in order to elicit a scenario, a piece of text must constitute a specific *partial description* of an element of the scenario itself' (1981: 129). These points and the structure of the examples in (32) lend support to our view, expressed already in Chapter 4, that effective staging, particularly thematisation, facilitates the processing of text. One function of thematisation at the text level may be to activate a particular scenario representation for the reader.

We should emphasise that Sanford & Garrod's claims relate to the ease or speed with which texts based on a coherent scenario can be processed. They do not suggest that texts for which a single scenario structure is not immediately available cannot be processed. Their scenario-based approach would encounter just as many problems as the frame-based approach if applied to the 'Pope meets Archbishop' text presented as (27) in Chapter 7. Their suggestion would no doubt be that such texts take longer to process.

Most of the textual material discussed by Sanford & Garrod is in the form of very brief constructed text which is designed for use in the controlled studies of the experimental psychology laboratory. In fact, this is a general feature of the 'texts' which appear in the work of psychologists investigating knowledge representation. Although Sanford & Garrod prefer the term 'scenario', they indicate that their notion of text-processing involving pre-existing knowledge representations has much in common with other studies

in which the term *schemata* is more generally used. If there is a difference between the use of these two terms, it appears to be that scenarios are situation-specific (At the Cinema; In a Restaurant), whereas schemata are much more general types of knowledge representations.

7.6.4 *Schemata*

We have already discussed one area of discourse studies, that related to story-grammars (cf. section 3.9), in which appeal was made to the existence of a particular type of *schema*. For the proponents of story-grammars, there exists a socio-culturally determined story-schema, which has a fixed conventional structure containing a fixed set of elements. One of these elements is the 'setting' and an initial sentence of a simple story (e.g. *All was quiet at the 701 Squadron base at Little Baxton*) can instantiate the setting element. It should be pointed out that, although a simple story may instantiate many elements in the story-schema, it is not suggested that the story has the schema. Rather, it is people who have schemata which they use to produce and comprehend simple stories, among many other things (e.g. place-descriptions in Brewer & Treyens (1981)).

Schemata are said to be 'higher-level complex (and even conventional or habitual) knowledge structures' (van Dijk, 1981: 141), which function as 'ideational scaffolding' (Anderson, 1977) in the organisation and interpretation of experience. In the strong view, schemata are considered to be deterministic, to predispose the experiencer to interpret his experience in a fixed way. We can think of racial prejudice, for example, as the manifestation of some fixed way of thinking about newly encountered individuals who are assigned undesirable attributes and motives on the basis of an existing schema for members of the race. There may also be deterministic schemata which we use when we are about to encounter certain types of discourse, as evidenced in the following conversational fragment.

(33)   A: There's a party political broadcast coming on – do you
want to watch it?
B: No – switch it off – I know what they're going to say
already.

However, the general view taken of schemata in the analysis of

discourse is much weaker. Rather than deterministic constraints on how we must interpret discourse, schemata can be seen as the organised background knowledge which leads us to *expect* or predict aspects in our interpretation of discourse. In fact, Tannen (1979: 138) uses the description 'structures of expectation' (adopted from Ross, 1975) to characterise the influence of schemata on our thinking. In Tannen (1980), there is also evidence that such expectations influence what type of discourse we produce. After watching a film (with no dialogue), a group of American subjects described in great detail the actual events of the film and what filming techniques had been employed. In contrast, a group of Greek subjects produced elaborate stories with additional events and detailed accounts of the motives and feelings of the characters in the film. Different cultural backgrounds can result in different schemata for the description of witnessed events.

This effect is not, however, caused by different cultural backgrounds alone. Anderson et al. (1977) presented a constructed text, partially repeated as (34), to a group of female students who were planning a career in music education and also to a group of male students from a weight-lifting class. Both groups had very similar cultural backgrounds, but would be predicted to have different 'interests'.

> Every Saturday night, four good friends get together. When Jerry, Mike, and Pat arrived, Karen was sitting in her living room writing some notes. She quickly gathered the cards and up to greet her friends at the door. They followed her ing room but as usual they couldn't agree on exactly Terry eventually took a stand and set things up. n to play. Karen's recorder filled the room t music. Early in the evening, Mike many diamonds . . .
> (Anderson et al., 1977: 372)

he discourse analysis
e female group with
sage as describing a
erson et al. found. They
group preferred an inter-
d some people playing cards
nderson et al. suggest that

people's personal histories, and interests (and sex, perhaps) contribute to the creation of 'higher-level schemata which cause them to "see" messages in certain ways' (1977: 377).

Both Tannen and Anderson derive their concept of 'schema' from the writings of Bartlett (1932). Bartlett believed that our memory for discourse was not based on straight reproduction, but was constructive. This constructive process uses information from the encountered discourse, together with knowledge from past experience related to the discourse at hand, to build a mental representation. That past experience, Bartlett argued, cannot be an accumulation of successive individuated events and experiences, it must be organised and made manageable – 'the past operates as an organised mass rather than as a group of elements each of which retains its specific character' (1932: 197). What gives structure to that organised mass is the schema, which Bartlett did not propose as a form of arrangement, but as something which remained 'active' and 'developing' (1932: 201). It is this 'active' feature which, combined with the experience of a particular piece of discourse, leads to the constructive processes in memory. The subject whom Bartlett (1932: 77) describes as remembering a story about 'two young men going down a river to hunt seals' in terms of 'two brothers going on a pilgrimage' has actively constructed the remembered discourse.

This 'active' aspect of Bartlett's proposed schemata is not generally a feature of other knowledge representations (e.g. frames) we have been considering. In some uses of the term 'schemata' by other writers, the 'active, developing' aspect is not promoted. For example, Rumelhart & Ortony propose that 'schemata represent stereotypes of concepts' (1977: 101). They present a schema for FACE which has subschemata for EYE, MOUTH, etc., which seems to have a lot in common with the slot and filler features of a frame. Their schema for FACE might best be described as a **prototype** for the various human objects called 'faces', in much the same way as Rosch (1973, 1977) and Rosch et al. (1976) suggest there are prototypic representations for natural and semantic categories like 'tree' and 'bird'. Viewed in this way, a schema is a fixed 'data structure'. Indeed, Rumelhart & Ortony propose schemata for linguistic knowledge which are very similar to the language frames of Minsky (1975). They propose that the GIVE schema has

discourse is much weaker. Rather than deterministic constraints on how we must interpret discourse, schemata can be seen as the organised background knowledge which leads us to *expect* or predict aspects in our interpretation of discourse. In fact, Tannen (1979: 138) uses the description 'structures of expectation' (adopted from Ross, 1975) to characterise the influence of schemata on our thinking. In Tannen (1980), there is also evidence that such expectations influence what type of discourse we produce. After watching a film (with no dialogue), a group of American subjects described in great detail the actual events of the film and what filming techniques had been employed. In contrast, a group of Greek subjects produced elaborate stories with additional events and detailed accounts of the motives and feelings of the characters in the film. Different cultural backgrounds can result in different schemata for the description of witnessed events.

This effect is not, however, caused by different cultural backgrounds alone. Anderson et al. (1977) presented a constructed text, partially repeated as (34), to a group of female students who were planning a career in music education and also to a group of male students from a weight-lifting class. Both groups had very similar cultural backgrounds, but would be predicted to have different 'interests'.

(34)   Every Saturday night, four good friends get together. When Jerry, Mike, and Pat arrived, Karen was sitting in her living room writing some notes. She quickly gathered the cards and stood up to greet her friends at the door. They followed her into the living room but as usual they couldn't agree on exactly what to play. Jerry eventually took a stand and set things up. Finally, they began to play. Karen's recorder filled the room with soft and pleasant music. Early in the evening, Mike noticed Pat's hand and the many diamonds . . .

(Anderson et al., 1977: 372)

The reader will no doubt have activated some discourse analysis 'schema' by now and have expectations that the female group with musical interests would interpret the passage as describing a musical evening. That is exactly what Anderson et al. found. They also found that the male, weight-lifting, group preferred an interpretation in which the passage described some people playing cards rather than musical instruments. Anderson et al. suggest that

people's personal histories, and interests (and sex, perhaps) contribute to the creation of 'higher-level schemata which cause them to "see" messages in certain ways' (1977: 377).

Both Tannen and Anderson derive their concept of 'schema' from the writings of Bartlett (1932). Bartlett believed that our memory for discourse was not based on straight reproduction, but was constructive. This constructive process uses information from the encountered discourse, together with knowledge from past experience related to the discourse at hand, to build a mental representation. That past experience, Bartlett argued, cannot be an accumulation of successive individuated events and experiences, it must be organised and made manageable – 'the past operates as an organised mass rather than as a group of elements each of which retains its specific character' (1932: 197). What gives structure to that organised mass is the schema, which Bartlett did not propose as a form of arrangement, but as something which remained 'active' and 'developing' (1932: 201). It is this 'active' feature which, combined with the experience of a particular piece of discourse, leads to the constructive processes in memory. The subject whom Bartlett (1932: 77) describes as remembering a story about 'two young men going down a river to hunt seals' in terms of 'two brothers going on a pilgrimage' has actively constructed the remembered discourse.

This 'active' aspect of Bartlett's proposed schemata is not generally a feature of other knowledge representations (e.g. frames) we have been considering. In some uses of the term 'schemata' by other writers, the 'active, developing' aspect is not promoted. For example, Rumelhart & Ortony propose that 'schemata represent stereotypes of concepts' (1977: 101). They present a schema for FACE which has subschemata for EYE, MOUTH, etc., which seems to have a lot in common with the slot and filler features of a frame. Their schema for FACE might best be described as a **prototype** for the various human objects called 'faces', in much the same way as Rosch (1973, 1977) and Rosch et al. (1976) suggest there are prototypic representations for natural and semantic categories like 'tree' and 'bird'. Viewed in this way, a schema is a fixed 'data structure'. Indeed, Rumelhart & Ortony propose schemata for linguistic knowledge which are very similar to the language frames of Minsky (1975). They propose that the GIVE schema has

three variables, a giver, a gift and a recipient, which are analogous to the 'cases' described by Fillmore (1968). They are clearly suggesting that schemata have fixed structures, containing set elements.

It may be, of course, that our background knowledge is organised and stored in some fixed schemata, together with some other, more flexible schematic structures. In whatever way they are represented, schemata seem to present the discourse analyst with one way of accounting for discourse production and interpretation which does not take place *ab initio* on each occasion. Like frames, scripts and scenarios, they are a means of representing that background knowledge which we all use, and assume others can use too, when we produce and interpret discourse.

The problems we noted with frames and scripts and scenarios are, however, also present for schematic representations. The selection and integration of schemata in the processing of a non-constructed piece of text such as (27) presents a substantial management problem. Given the proposals in the literature for how knowledge may be represented, future research must be aimed at devising heuristics for the *selection*, on a particular occasion, of the *relevant* partial representation (and no more) that is required for the local interpretation of discourse fragments. In so doing, this research will necessarily also have to devise controls on stereotypic knowledge representations which allow them to recognise 'weird' events which nevertheless fit the stereotype format. If an understander system decides that John ate a steak after reading the following text, then it has failed to 'understand' what most human processors understand about this particular restaurant scenario.

(35)   John is pretty crazy, and sometimes does strange things. Yesterday he went to Sardi's for dinner. He sat down, examined the menu, ordered a steak, and got up and left.
                                                        (from Kaplan, 1981: 131)

7.6.5   *Mental models*

A view of how we interpret discourse (and experience) which does not appeal to stereotypic knowledge or fixed storage systems has been put forward by Johnson-Laird in a series of papers. Johnson-Laird (1981a) argues against an approach to the meaning of sentences which depends on a decomposition of word-

meaning having to take place. An example of a decomposition view is that of Katz & Fodor (1963) where the 'meaning' of *man* is decomposed into *human, adult, male*. The conceptual dependency type of analysis used by Schank (1972), discussed earlier, is another example. Johnson-Laird proposes that we are indeed capable of decomposing word-meaning, but that we do not typically do so in our normal understanding of sentences. He suggests that a sentence like (36) receives an immediate interpretation which makes sense to most people as praise for the book.

(36)   This book fills a much needed gap.

Upon further analysis, however, we can work out that the sentence is actually saying that it is the gap, not the book, which is needed. To account for this everyday non-analytic process of comprehension, Johnson-Laird proposes that we use words in a sentence as 'cues to build a familiar mental model' (1981a: 122). A mental model is a representation in the form of an internal model of the state of affairs characterised by the sentence. We should note that although such models are not described as stereotypic, the term 'familiar' is rather smuggled into the description without any account of what 'being familiar' is based on. There are, moreover, theoretical problems with the concept of an 'internal' model, which Johnson-Laird (1981a: 117) acknowledges. However, he notes that the experimental evidence on instantiation (cf. Anderson & Ortony, 1975; Anderson et al., 1976; Garnham, 1979) supports a view of understanding via mental models, rather than via the decomposition of word meaning. When subjects were asked to recall a sentence like (37), Anderson et al., found that the word *shark* was a much better recall cue than the word *fish*.

(37)   The fish attacked the swimmer.

Johnson-Laird accounts for this finding by suggesting that readers interpreted the sentence by constructing a mental model in which the relevant event and entities were represented. We should note that this is, at least, a text-specific model, since it is very easy to imagine texts in which the term *fish* would not bring *shark* to mind at all.

Johnson-Laird (1980, 1981b) specifically appeals to the ideas of model-theoretic semantics in support of his notion of mental

models. In formal semantics, a model structure can be used to represent a possible state of affairs at a particular point in time and space which can correspond to the 'meaning' of a sentence (cf. Thomason (ed.), 1974; Partee (ed.), 1976). We shall not describe formal model-theory in any greater detail here, except to point out that it is not intended as a psychological account of meaning or understanding. As Johnson-Laird observes, model-theory relates language to the world, but not by way of the human mind. What a psychologically interesting model-theory has to be concerned with is that 'in so far as natural language relates to the world, it does so through the mind's innate ability to construct models of reality' (Johnson-Laird, 1981b: 141). These *models* of reality are, of course, representations of the way the world is. They may differ from one individual to the next. This is unavoidably the case when such models are the result of a listener's (or reader's) comprehension of discourse. According to Johnson-Laird (1981b: 139):

a major function of language is to enable one person to have another's experience of the world by proxy: instead of a direct apprehension of a state of affairs, the listener constructs a model of them based on a speaker's remarks.

As a simple example, Johnson-Laird & Garnham (1979) point out that the interpretation of a definite description is not determined by uniqueness in the world, but uniqueness in the local model constructed for the particular discourse. If a speaker says:

(38)   The man who lives next door drives to work.

the hearer may have a model of a particular state of affairs in which there is an individual (neighbour of speaker, has a car, has a job, etc.), but the hearer is unlikely to assume that the speaker has only one neighbour.

The proposal that understanding takes place via the construction of mental models leads Johnson-Laird to a view of comprehension and inference which is quite different from those we have already investigated. In this view, there is a level of comprehension which is based on the construction of an initial mental model which, as we noted with example (36), need not result from any elaborate consideration of the text encountered. There are, however, other levels of comprehension which result from the manipulation of the mental model constructed and which can lead to the abandonment

of the initial model and the construction of another. In this process of manipulation, there are no rules of inference, there are only procedures for testing the constructed mental model to find out if it fits the state of affairs described by the text. As an illustration of this process, Johnson-Laird (1980) takes an example (39) of the type used in discussions of syllogistic inference.

(39)   All of the singers are professors.
         All of the poets are professors.

Given the pair of premises in (39), we can construct a model with, for instance, six individuals in a room and assign the roles of singer, poet and professor to those individuals in a way that fits the state of affairs described by the two sentences in (39). One model which immediately comes to mind is that, for all six individuals, the following representation (40) is true:

(40)   singer = professor = poet

According to this model, the conclusion that *all of the singers are poets* or *all of the poets are singers* is justified. Johnson-Laird & Steedman (1978) report that, for many people, this conclusion is the natural one. It is possible to test the model in (40) against the state of affairs described in (39) and find that it is not necessarily a correct representation. By manipulating the model, it is possible to arrive at a representation (41) in which a is true for three individuals and b is true for the other three.

(41)   a. singer = professor

         b. poet = professor

On the basis of (41), one might conclude that *none of the singers are poets*. On further manipulation, one might arrive at a model (42) in which a is true for four individuals, b is true for one, and c is true for the other one.

(42)   a. singer = professor = poet

         b. singer = professor

         c. poet = professor

So, one could conclude that *some of the singers are poets*.

It should be clear that the sentences in (39) can give rise to several different versions of a mental model involving the six individuals with different identities. The process of manipulation of the model which has just been described is characterised by Johnson-Laird (1980: 81) as 'testing your mental model to destruction'. The discourse analyst may not be as interested as the logician in carrying out the 'testing' procedure to its extreme, but he must acknowledge that Johnson-Laird's notion of understanding via the construction and manipulation of mental models provides a useful metaphor for the way a piece of text can be 'understood' at different levels. It also accommodates that aspect of discourse understanding (which we have argued for already in section 6.2) which allows interpretations in different receivers' minds to differ from the interpretation intended by the discourse producer. The individual hearer's mental model of the discourse can differ from the speaker's, and there is no suggestion that *the text* is, in any sense, *the model*.

It should be apparent from the consideration of the sentences in (39) how Johnson-Laird intends us to understand his claim that, in the mental model approach to understanding, there are no rules of inference. Whereas the formulae in (40), (41) and (42) are normally considered inferences from (39), in Johnson-Laird's analysis they are different versions of a mental model for the text. That is, what we normally describe as a process of inferring one state of affairs on the basis of another is presented in this alternative view as building a model of one state of affairs, or building another model from another state of affairs. From a discourse analyst's point of view, this distinction is of little practical significance.

Johnson-Laird's view of discourse understanding via mental models is never described in terms of the sets of stereotypical elements found in 'frames' or the sets of characteristic events of a narrative 'schema'. Possibly for this reason, the practical details of mental models remain elusive. They seem to represent a way of thinking about how we understand discourse rather than a way of doing analysis of discourse. Yet the problem we have frequently noted with other methods of representing discourse processing and understanding – that of fixing the constraints on what knowledge we use – must also exist for mental models. When we construct a mental model for a piece of discourse, we use some of our

pre-existing knowledge and experience to get a 'picture' of the state of affairs described by the discourse. How is it that we do not use *all* of our pre-existing knowledge? Putting this question in more specific terms, will a mental model theory predict that, in asking subjects to recall a sentence like *The fish attacked the swimmer*, not only is *shark* a better cue than *fish*, but that *blood* or *teeth* or *ocean* or *bite* or *splash* are also better? At the moment, we have no answers to these questions.

As it is presently described, the theory of mental models actually predicts massively detailed mental representations of any event encountered, whether in life or via text. Admittedly, one of the advantages of the concept of a mental model is that it allows for a richer representation than the rather bare outlines of the stereotypic versions found in scripts and scenarios. The scenario example, quoted earlier as (32), to demonstrate the 'In court – the lawyer' connection, seems to describe a strangely empty and non-detailed court-scene which is at odds with the experience of most people. However, the unconstrained potential of the mental model concept takes us to the other extreme. It would lead to a pathological inability to process text at all. A well-documented case-history of an individual whose 'mental models' were unconstrained is presented in Luria (1969). The incapacitating effects of this lack of constraints can be detected in the following account:

Last year I was read an assignment having to do with a merchant who had sold so many meters of fabric . . . As soon as I heard the words *merchant* and *sold*, I saw both the shop and the storekeeper, who was standing behind the counter with only the upper part of his body visible to me. He was dealing with a factory representative. Standing at the door of the shop I could see the buyer, whose back was toward me. When he moved off a little to the left, I saw not only the factory but also some account books – details that had nothing to do with the assignment. So I couldn't get the gist of the story.

(Luria, 1969: 66)

The outstanding problem for Luria's patient, and also for the discourse analyst who wishes to represent the interaction between previous knowledge / experience and the comprehension of the discourse at hand, is to reach a working compromise. In this compromise representation, there should be enough richness of detail to capture the potential complexity of our pre-existing knowledge / experience, but there should also be a constraint on

how much of this richness of detail we actually use in our processing of the discourse we encounter.

7.7 Determining the inferences to be made

Much of the data presented in this chapter is of the type that has generally been treated as requiring **inferences** on the reader's part to arrive at an interpretation. The rather general notion of inference appealed to is used to describe that process which the reader (hearer) must go through to get from the literal meaning of what is written (or said) to what the writer (speaker) intended to convey. For example, the general view of the interpretation of an utterance such as (43) – used to convey an indirect request – is that the hearer works from the literal meaning to a meaning like (43a) via inference(s) of what the speaker intended to convey.

(43)   It's really cold in here with that window open.

(43a)   Please close the window.

In other words, utterance (43) does not 'mean' (43a). Rather, the hearer, on receiving (43) in a particular context, must infer that the speaker intended it to convey (43a). As evidence that some inferential process is required in the interpretation of indirect requests, Clark & Lucy (1975) demonstrated that, across a wide range of indirect versus direct forms, readers performing a verification task consistently took longer with the indirect forms. The additional time taken, Clark (1978) claims, is required by the reader's inferential processing of the indirect request.

Very similar evidence is presented by Haviland & Clark (1974) to show that 'identifying referents for definite noun phrases is a highly inferential activity' (Clark, 1978: 313). Haviland & Clark found that determining the referent for *the beer* in (45b) took readers significantly longer than in (44b).

(44)   a. Mary got some beer out of the car.
         b. The beer was warm.

(45)   a. Mary got some picnic supplies out of the car.
         b. The beer was warm.

This finding is explained in terms of a particular aspect of the

inferential process described as forming a **bridging assumption**. The bridging assumption required between (45a) and (45b) is that shown in (45c).

(45)   c. The picnic supplies mentioned include some beer.

Forming this type of bridging assumption takes time and so the difference in comprehension times noted between (44b) and (45b) is accounted for. The implication from this type of research finding is that inferences take time.

7.8 Inferences as missing links

The information in (45c) can be seen, in formal terms, as the **missing link** which is required to make an explicit connection between (45a) and (45b). Is it possible, then, to think of an inference as a process of filling in the missing link(s) between two utterances? This seems to be implicit in the research of Clark and his co-authors and also seems to be the basis of Prince's (1981) category of 'inferrable', described already in section 5.3.2. Indeed, there are many examples in the literature concerning definite descriptions which we could treat in terms of the 'missing link' phenomenon. Let us consider some of these examples, which we will present with the a and b sentences ('the text') as *linked* via the information in the c sentence ('the missing link').

(46)   a. I bought a bicycle yesterday.
         b. The frame is extra large.

                                              (Chafe, 1972)

         c. *The bicycle has a frame.*

(47)   a. I looked into the room.
         b. The ceiling was very high.

                                              (Clark, 1977)

         c. *The room has a ceiling.*

(48)   a. This afternoon a strange man came to my office.
         b. His nose was nearly purple.

                                              (van Dijk, 1977)

         c. *The man has a nose.*

(49)   a. I got on a bus yesterday
         b. and the driver was drunk.

                                              (Prince, 1981)

         c. *The bus has a driver.*

In each of these examples, the missing link expresses a type of generally true relationship which might take the form of a universally quantified proposition such as *Every X has a Y*. In fact, each of the four *c* sentences in (46) – (49) expresses information which we might expect to be represented in one of the stereotypic knowledge formats (e.g. frames, schemata) discussed already in section 7.6. The same could be said for the relationship (*Every X is a Y*) expressed in the *c* sentences of the following two examples.

(50)   a. A bus came roaring round the corner.
        b. The vehicle nearly flattened a pedestrian.
                            (Garrod & Sanford, 1977)
        c. *The bus is a vehicle.*

(51)   a. Draw a diameter in black.
        b. The line is about three inches.
                                 (Yule, 1981)
        c. *The diameter is a line.*

These types of 'generally true' missing links are also presented in terms of a connection between the verb of one sentence or clause, and the definite noun phrase of another, as in the following examples.

(52)   a. She decided to sell the cow
        b. and buy a shop with the money
                                (Chafe, 1972)
        c. *Selling involves money.*

(53)   a. It was dark and stormy the night the millionaire was murdered.
        b. The killer left no clues for the police to trace.
                          (Carpenter & Just, 1977b)
        c. *Murdering involves a killer.*

(54)   a. Mary dressed the baby.
        b. The clothes were made of pink wool.
                          (Sanford & Garrod 1981)
        c. *Dressing involves clothes.*

This last example (54) was used in a controlled experiment by Sanford & Garrod to test whether the type of missing link involved required the additional processing time which Haviland & Clark

(1974) noted in connection with the *picnic supplies–beer* example, quoted earlier in this chapter as (45). When the times taken to understand the *b* sentence in (54) were compared with those for the *b* sentence in (55), no significant differences was found.

(55)   a. Mary put the baby's clothes on.
        b. The clothes were made of pink wool.

In other words, despite the fact that we can point to a missing link in (54c), the experimental subjects did not behave as if that missing link required additional processing time to work out. Does this result nullify the finding of Haviland & Clark (1974) that the existence of a missing link creates additional processing requirements? Sanford & Garrod do not think so. They suggest that when the missing link is already part of the knowledge representation (e.g. frame, schema) activated by one part of the text, no additional processing is required to understand subsequent reference to another element in that knowledge representation. They claim that because *dressing* activates *clothes* in our representation of the first part of the text (54a), subsequent mention of *the clothes* is understood as quickly as it would be if *the clothes* had already been explicitly mentioned, as in (55a). However, since *picnic supplies* did not automatically activate beer in the knowledge representations of Haviland & Clark's subjects, they had to make a bridging assumption and so took additional processing time.

It seems, then, that we have (at least) two categories of missing link. One kind is automatically made and does not result in additional processing time and the other is not automatic, but is the result of a bridging assumption and leads to additional processing time. If we wish to maintain, as was suggested earlier, that inferences take time, then it should follow that those missing links which are automatically made (and do not take additional processing time) are not to be described as inferences. This would be the natural conclusion of any researcher who, working on an empirical basis, finds no evidence for the existence of a hypothesised process. Let us assume, then, that 'missing links' are formally identifiable sentences which can be shown to provide a connection, in formal cohesive terms, between text sentences. Providing missing links may be part of an exercise in text-representation, but that is not the same as providing a representation of what people are doing in comprehending text. We could then draw a distinction between

inferences and missing links in the following terms: texts may have formal missing links, but it is readers and hearers who make inferences. Identifying missing links is not the same as identifying inferences.

7.9 Inferences as non-automatic connections

Sanford & Garrod's proposal that automatic connections are made between elements in a text via pre-existing knowledge representations could be used as a basis for deciding which missing links are, and which are not, likely to be inferences. That is, all the *c* sentences in (46)–(54) are automatic connections, and consequently should not count as inferences, but the connection between *picnic supplies* and *beer* in (45) is non-automatic and ought, therefore, to be treated as an inference. Such a proposal appears to be in line with de Beaugrande's suggestion that there is a process, in our understanding of what we read and hear, of 'spreading activation' which 'results naturally from concept activation in ideation and comprehension without specifically directed impulses' (1980: 229). Those 'specifically directed impulses', on the other hand, are expressly aimed towards overcoming discontinuities or gaps in the reader's (hearer's) understanding of what he reads (hears) and are more properly treated as inferences. This distinction allows us to think of non-automatic connections (inferences) as requiring *more interpretive work* on the reader's (hearer's) part than automatic connections made via pre-existing knowledge.

The idea of 'automatic connections' can also be usefully applied to an aspect of text understanding which has been discussed in terms of 'informational inferences' (Warren et al., 1979). Since the type of 'information' described appears to involve automatic connections across text sentences, it may be that the phenomenon has been inappropriately characterised as an example of 'inference'. Warren et al. (1979) claim that, in our understanding of a text, we continually need to know the answers to a set of *who, what, where* and *when* questions. Arriving at the answers to these questions, at a particular point in a text, is accomplished, they suggest, by making 'informational inferences'. Thus, on encountering the final sentence, *he tied her shoelaces together*, in the text shown here as (56), the reader has to infer who is doing what to whom, where and when.

(56)   It was Friday afternoon.
        Carol was drawing a picture in the classroom.
        David felt mischievous.
        David decided to tease Carol.
        When Carol was not looking,
        he tied her shoelaces together.

(Warren et al., 1979: 24)

It may be particularly unfortunate that Warren et al., choose to discuss 'informational inferences' in relation to our understanding of such a simple piece of text. Given the principles of analogy and local interpretation which we described in Chapter 2, there is a fairly automatic understanding of who is doing what to whom, when and where, in the final sentence of this text. Since there is no competition between different times, different locations or different referents, the reader has very little interpretive 'work' to carry out in understanding the final sentence. Let us assume that the reader's understanding that *David tied Carol's shoelaces together in the classroom on Friday afternoon* is a result of making fairly automatic connections and is not the product of any inference-making at all.

There are, however, some texts which, for some readers, will pose more substantial comprehension problems of the *who, what, where* and *when* variety than the simple text in (56). We shall consider this issue in relation to examples (61) and (62) later.

Warren et al. continue the text of (56) with the sentence shown in (56a). They suggest that a 'logical inference' has to be made to connect the final sentence of (56) with the sentence in (56a).

(56a)   Carol tripped and fell down.

This type of 'logical inference', alternatively described as an 'enabling inference' by Hildyard & Olson (1978), is typically supplied by readers to make a connection in terms of action A causing action B. Interestingly, Warren et al. describe the 'causation' relationship in their example in terms of a 'specific prediction' (1979: 26) which the reader of (56) is likely to make. If a 'logical inference' of this type can be based on a prediction, then it is clearly in the category of automatic connections. Presumably the knowledge-base used in making such predictions is concerned with general cause and effect relationships. This type of knowledge will lead the reader not to derive a 'logical inference' to connect the first two sentences of (56). That is, the fact that it was *Friday afternoon*

did not cause *Carol* to start *drawing a picture*. In the simple text under consideration, the notion of 'logical inference' seems to lead to automatic connections. However, there are texts in which a causation relation may, in fact, be far from automatically made. This is because the reader may have to ask a *why* or *how* question with regard to some action or event described in a text. Such questions also give rise to what Warren et al. wish to describe as 'elaborative' and 'evaluative' inferences. At this point, the categories of the inference types proposed in the taxonomy begin to merge into one another. We shall try to illustrate 'elaborative' and 'evaluative' inferences in the discussion of extract (61) later in this chapter.

For the moment, we shall concentrate on the implications of an approach which maintains that automatic connections made in text comprehension should not be treated as inferences.

One of the simplifying assumptions made in many psycholinguistic investigations of text understanding is that the experimental subjects are a representative sample of a population which has fairly homogeneous background knowledge and experience. Another assumption is that the two-sentence text, specially constructed and decontextualised, is a representative sample of the linguistic material encountered by the language-user as naturally occurring discourse. On the basis of these two assumptions, it is possible to draw a distinction between the processing of texts which contain automatic connections (*dressing the baby – the clothes*) and those which contain non-automatic connections (*picnic supplies – the beer*). We can then suggest that only the latter type should be treated as an example of inference, because we have evidence (additional time taken) that the reader has had to undertake some additional interpretive 'work' in his processing of the text. This is basically a useful distinction and may provide a general heuristic for predictions about which texts will probably be more difficult to process than others.

The danger of this approach, however, is that it tends to identify inferences with specific text-connections and to base those text-connections on the words in the text. Consider again the idea that, if an element is activated because it is necessarily part of the reader's (hearer's) pre-existing knowledge representation, then it receives 'direct interpretation' (Sanford & Garrod, 1981: 105), and does not require additional processing time. Now consider Haviland & Clark

presenting their *beer – beer* (44) and *picnic supplies – beer* (45) examples to a group of real ale enthusiasts who often indulge their enthusiasm on picnics at the local park. By Sanford & Garrod's prediction, there should not be, for this group, any differences in processing time under the two conditions. This would also be predicted by Anderson et al.'s (1977) concept of schema, described already in connection with extract (34). What this means is that the identification of a connection as 'automatic' or 'non-automatic' cannot be made independently of the person(s) considering the text. For some people, *beer* is an automatic component of *picnic supplies*, for others it has to be included on a particular occasion because understanding the text at hand requires its inclusion.

A second problem arises in connection with determining exactly which elements will be automatically activated via the reader's (hearer's) pre-existing knowledge representations. Given the following sentence (57), we presumably should be ready to make a 'direct interpretation' of the elements referred to by some of the definite expressions in the sentences listed under (58).

(57)   Socrates is a lovely striker of the ball.

(58)   a. His height gives him a great advantage.
        b. His father was in love with Greek culture.
        c. The Brazilian midfield man is interested in playing in Europe.
        d. The goalkeeper didn't even have time to move.
        e. The nail on the index finger of his left hand is broken.

The first point to be made is that, for many reasons, some of these potential co-text sentences in (58) may not be interpretable at all without the general context of (57). If that is the case, then knowledge-activation is clearly context-dependent for naturally occurring texts. Example (57) is quoted from a commentary on a soccer match during the World Cup Finals in Spain, in June 1982. The sentence which actually follows (57) in the commentary is (58d). The definite expression *the goalkeeper* may, of course, be quite automatically interpreted given the hearer's activated knowledge of elements in his soccer match 'frame'. Notice that this 'automatic' connection is not made across the two-sentence text formed by (57) and (58d) alone. Sentences (58a–c) are taken from other parts of the commentary, but all have definite expressions

which depend, for their interpretation, on a connection to the 'Socrates' of sentence (57). Perhaps the most obvious connection is from 'Socrates' to 'his height', but even this connection is hardly automatic in this text without some additional connections which make Socrates a soccer player who hits the ball with his head, on occasion, hence the advantage of 'his height'. The connection between 'Socrates' and 'his father' might seem relatively simple, since every person has a father. Yet, in this text, the mention of his father is embedded in what seems to be an explanation for this particular soccer player having the name he does. The connection between 'Socrates' and 'his father', in this text, may require the reader to 'fill in' several other connections, none of which is necessarily derived from the activated soccer match 'frame'. The connection between (57) and (58c) is of a type which is frequently made in sports and news reports, and we have discussed this role-related aspect of reference already in Chapter 6. Whether this type of connection is automatic or not clearly depends on very localised knowledge, because it is not of the same generality as the 'every bus is a vehicle' type noted in example (50) earlier. Finally, sentence (58e), which is not taken from the commentary, but is a constructed sentence, is presented as an example of a definite expression used to refer to an element which is a necessary part of any person. Every person has a 'nail on the index finger of his left hand', but would we really expect this information to be automatically activated by the mention of a person's name in a preceding co-text sentence? If the answer to this question is 'yes', then what human feature is not activated? The problem is very similar to those noted with the representation of context in Chapter 2 and with representing background knowledge in section 7.6 – how do we set the boundaries on these representations? The example in (58e) is presented as part of what could be described as a *reductio* argument against the unconstrained nature of the knowledge representations which are claimed to provide automatic connections within texts. Maratsos (1971) makes a similar type of argument regarding the use of definite noun phrases. Some connections appear to be automatic, as exemplified in examples (46)–(54), yet others, though clearly filled in via aspects of our knowledge representations, as between (57) and (58a–e), are not automatic for the majority of readers (hearers).

A third problem with the automatic connection via background-knowledge view is the assumption that the connection can be described in terms of a decomposition of lexical meaning. Chafe (1972: 61) suggests that this may be a reasonable approach and Sanford & Garrod make the point in processing terms: 'when a verb like *dress* is encountered, this will evoke from memory a representation which contains slots for a variety of entities implied in the meaning of the verb, such as *clothing*' (1981: 108). If this really were the case, then there would be an extremely large, and massively redundant, representation which would be unlikely to lead to the automatic connection type of processing indicated in their experimental findings. Why would *clothing*, for example, enter into the representation of our understanding the following two constructed texts?

(59)   a. Mary dressed the baby's arm.
         b. The bandage was made of white cotton.

(60)   a. Mary dressed the turkey.
         b. The entrails spilled out into the bowl.

It is clearly not the lexical item *dress* alone which is the source of the activated knowledge representations we use in the comprehension of two-sentence texts such as (59), (60) and (54).

Given these problems, it may be that the discourse analyst can make only very limited use of the results of psycholinguistic experiments on the nature of inference. The two-sentence text, specially constructed and presented in isolation from communicative context, is not generally what the discourse analyst encounters as data, nor what the language-user encounters as a linguistic message. The controlled experiment offers insight into some aspects of our processing of sentences, but it can be misleading to take discourse processing as generally occurring in this concentrated and narrowly delimited way.

7.10. Inferences as filling in gaps or discontinuities in interpretation

We have argued against equating inferences with any form of connection between sentences in a *text*. We have emphasised that inferences are connections people make when attempting to reach an interpretation of what they read or hear. We have also

suggested that the more interpretive 'work' the reader (hearer) has to undertake in arriving at a reasonable interpretation of what the writer (speaker) intended to convey, the more likely it is that there are inferences being made. The problem with this view is that it leaves 'inferencing' as a process which is context-dependent, text-specific and located in the individual reader (hearer).

While we believe that this is a correct view and that it is, in principle, impossible to predict the *actual* inferences a reader will make in arriving at an interpretation of a text, we may be able to make predictions regarding particular aspects of individual texts which readers will generally have to interpret on the basis of inference. Such predictions will be closely related to some concept of 'depth of processing'. Clearly, the reader who casually skims across the news article presented below as (61) while sitting in the dentist's waiting room, is likely to be 'reading' the text in a qualitatively different way from the reader who is anticipating being asked comprehension questions after he has finished the text. Since the type of 'understanding' normally discussed in discourse analysis, in psycholinguistics, and in computational modelling, tends to be of the latter type, let us consider the text in terms of a set of comprehension questions which might be asked of the reader. If answering some of these questions appears to involve the reader in additional 'work' such as filling in gaps or discontinuities in his interpretation, then we may find a basis for predicting what kind of inferences will be required.

(61)   1. The agents of the Public Security Bureau seemed intent on terrorizing their victim, and they succeeded.
         2. It was 1 a.m. when they marched into Peking's sprawling Friendship Hotel where many foreigners working in China live.
         3. The police told room clerks to awaken American teacher Lisa Wichser, 29, and tell her that an urgent telegram had arrived for her.
         4. When the petite, sandy-haired and somewhat sleepy Wichser appeared to claim it, she was handcuffed and hustled without explanation into a police car.
         5. Technically, at least, the graduate student from Noblesville, Ind., had not been arrested.

                                                       (*Time*, 14 June 1982)

If we first try to answer the set of *who, what, where* and *when*

questions, proposed by Warren et al. (1979), we should arrive at a partial representation of what we understand about the persons and events described in this text. The first thing we may note is that there isn't the simple proper name–pronoun connection throughout, as there was in (56). Instead, there is an array of different definite descriptions. We are not explicitly told that *the agents of the Public Security Bureau* are the same people as *the police* and that they *handcuffed* an individual. Nor are we explicitly informed that the expressions *their victim, American teacher Lisa Wichser, 29, the petite, sandy-haired and somewhat sleepy Wichser* and *the graduate student from Noblesville, Ind.* are all being used to refer to this particular individual. Unless the reader has some previous, specialised knowledge about this news item, he most likely has to 'work out' that *the police* in line 3 are the same individuals, more or less, as *the agents* in line 1. Some comparable interpretive 'work' has to be involved in equating *their victim* with *Lisa Wichser* and then with *the graduate student*. The interesting thing about this last expression is that it is a definite expression apparently being used to refer to an individual already introduced into the discourse domain and so a candidate for 'given' status. However, the information carried by the expression is 'new', in the discourse. It is, as we have noted in Chapter 5, a 'given' entity in a 'new' role. We suggest that, unless the reader has specialised knowledge about the entity in the mentioned role, this type of expression will create a potential discontinuity in the reader's interpretation and require inferencing.

Perhaps this last point can be more forcefully made by considering a brief text in which highly specialised knowledge is assumed and within which totally mistaken connections could be inferred by the uninformed reader.

(62)   As bullion levels dropped below the psychological $300 barrier, putting most high-cost mines into loss, kaffirs fell sharply, with 'the heavies' closing $1 to $4 down.

                                                  (*The Guardian*, 22 June 1982)

One might, on reading (62), infer that *kaffirs* are *bullion levels* or *high-cost mines*, or that '*the heavies*' are *high-cost mines* or 'bullion dealers' or some types of metals. We have been reliably informed that none of these inferences is justified, in fact.

Returning to extract (61), we can note that the *when* and *where*

of the described events are only mentioned explicitly in sentence 2, but that we can operate with the 'no-spatio-temporal-change-unless-indicated' principle, expounded in Chapter 2, to place the events described in the other sentences in the same spatio-temporal location. However, in order to answer the question – *where was Lisa Wichser sleeping?* – some readers may feel that they have to perform some interpretive 'work'. Other readers may answer the question without hesitation and feel that no inferences have to be made. Clearly, it is not stated explicitly in the text that *Lisa Wichser* is even living in the *Friendship Hotel*. In order to answer the question, we would tentatively suggest, the reader would probably have to fill in the discontinuity existing in his interpretation. Such a conclusion, however, is intended largely as a hypothesis which might be tested in some experimental investigation with 'real' data such as the text in (61). At the moment, we can only suggest likely points at which inferences may be required.

Once one goes beyond the strictly factual considerations of *who, what, where* and *when* questions, the need for inference becomes very obvious. If *how* and *why* questions are asked, we immediately have to make what Warren et al. (1979) describe as 'elaborative' and 'evaluative' inferences. An elaborative inference would involve, for example, deciding how *Lisa Wichser* was probably dressed when she appeared to receive her *telegram*. An evaluative inference might involve deciding whether the police behaviour was justified or whether *the telegram* actually existed. It might be made in response to a question about why *Lisa Wichser* was *handcuffed* and taken away. A large part of our comprehension of what we read and hear (and see, no doubt) is, after all, a product of our making sense of the motivations, goals, plans and reasons of participants in described or witnessed events. Evaluative inferences must clearly be based on more than the reader's interpretation of the literal description of events in the text. They might be based on such diverse beliefs that, on the one hand, all Americans in China are CIA agents or, alternatively, that the Chinese continually harass foreigners for no reason. Such inferences will readily be made by a reader to try to account for behaviour which is described, but not explained, in a text. They represent the open-ended aspect of 'filling gaps' in text-described events which a reader may perform in arriving at his or her 'comprehension' of a text.

Given this 'open-ended' feature of inferencing, it is extremely difficult to provide, for any naturally occurring text, the single set of inferences which an individual reader has made in arriving at an interpretation. One might say, as Clark (1977) does, that there is a set of *necessary* inferences which every reader must make to arrive at an interpretation. However, those necessary inferences appear to be exactly the type which, on existing experimental evidence, do not require additional processing time. The fact that the *room clerks*, mentioned in sentence 3 of (61), must work in the *Hotel*, mentioned in sentence 2, would have to be treated as an automatic connection and likely to produce no evidence (in empirical terms) of processing via inference. The discourse analyst may consequently find himself in the confusing position that the so-called 'necessary' inferences may not justifiably be described as inferences at all, and the 'elaborative' and 'evaluative' inferences may be, in principle, undeterminable. In other words, the analyst may be left with no secure basis for talking, in analytic as opposed to intuitive terms, about the inferences involved in the comprehension of texts.

This rather bleak conclusion is not intended as a suggestion that the nature of inference is beyond description. Rather, it is an attempt to state the existing problem quite specifically. The illusion that we can determine the nature of inference by inventing a taxonomy and illustrating each type with a constructed set of sentences, as in Warren et al. (1979) and Clark (1977), is exposed whenever a naturally occurring piece of text is encountered (see van Dijk (1981) for a criticism of this taxonomic approach). The fact is that, until we can develop experimental techniques which allow us to draw conclusions about how people process naturally occurring discourse in 'real-life' contexts, we shall continue to underdetermine human understanding and overindulge our simplistic analytic metaphors. This applies not only to the nature of inference, but to the more general concept of comprehension itself.

At the present time, the most we can say is that a highly cohesive text which has few 'missing links' will require a lot of space to convey very little information, but will not demand a lot of interpretive 'work', via inference, on the part of the reader. However, it is typically the case that the texts which a reader will normally encounter will show a minimal amount of formal cohesion, assume massive amounts of existing background knowledge,

and normally require the reader to make whatever inferences he feels willing to work for in order to reach an understanding of what is being conveyed. As an extreme example of this latter type of text, we leave the reader with extract (63) and ask him / her to try to write out even a few of the connections (one might say 'inferences') which have to be made in order to produce a coherent interpretation for what the reader thinks the text-producer intended to convey.

(63)   *Swap a child this summer: Family Centre Special Education Centre*

When 'O' or 'A' levels loom, there aren't many subjects in which parents can give direct help: except languages. The only satisfactory way to learn a language is to be immersed in it for a while. And since just on the other side of the water, a European teenager is in the same position with his English as yours with his French or German, a swap seems obvious. Three weeks or so in each other's family and the candidates surely will have that part of the G.C.E. or *bac* safely buttoned up. It's a simple idea and often it works very well but many mistakes are made by attempting it too soon. However, a well-adjusted child of 14+ should be able to cope.

(*Good Housekeeping Magazine*, 14 April 1976)

7.11   **Conclusion**

In this book we have tried to assemble some of the ingredients which would be required to construct an account of how people use language to communicate with each other. We have paid particular attention to ingredients which are dominant in the literature. We have tried to show that, at the present time, workers in discourse analysis have only a partial understanding of even the most-studied ingredients. There is a dangerous tendency, among established scholars as among students, to hope that a particular line of approach will yield 'the truth' about a problem. It is very easy to make claims which are too general and too strong. We have tried to show that some of the established wisdom in the area of discourse analysis may illuminate some aspects of discourse processing and of language use, but that all approaches open up yet more gaps in our understanding.

We have only discussed some of the relevant questions. We have largely ignored many aspects of the language of discourse which

receives attention in mainstream linguistics. We have concentrated on questions relating to reference and to the general issues of coherence and relevance. We have left virtually untouched several areas which occupy scholars working on the interaction of semantics and syntax – questions of aspect, tense, modality, quantification, negation, adverbial modification and so on, as well as relevant issues like the influence of metaphor in the interpretation of discourse.

Such an approach obviously has pitfalls. We hope that the losses, in terms of the occasional simplified explication, will be outweighed by the gains in terms of accessibility. Above all, we hope that the analysis of discourse, undertaken in the manner presented in this book, will not only provide the reader with insights into the workings of his own language, but also encourage him to think afresh about the nature of that complex cognitive and social phenomenon we call 'discourse'.