# Chapter 4
# The Generation of Messages

The construction of a preverbal message is a first step in the generation of speech. This step is usually initiated by the conception of some communicative intention; the speaker wants to achieve some purpose by means of saying something, and he wants the addressee to recognize that intention from what is said. Given the communicative intention, the speaker will select information for expression that is expected to be instrumental in realizing the goal. This information should make it possible for the cooperative addressee to infer the intention. These issues are discussed in section 4.1.

But information can be instrumental only if it is relevant to the situation of interaction, and that situation is a continuously changing one. As discourse proceeds, intermediary goals are achieved or blocked, new relevant facts appear, and so forth. A major task of the speaker while constructing messages for expression is to keep track of what is happening in the discourse situation. This "bookkeeping" is the topic of section 4.2.

We will then proceed to message construction proper. It will be suggested that the preparation of a message involves two steps. The first one will be called *macroplanning*. It consists in the elaborating of the communicative intention as a sequence of subgoals and the selection of information to be expressed (asserted, questioned, etc.) in order to realize these communicative goals. This determines the content of the subsequent speech acts. The second step, *microplanning*, is concerned with the further shaping of each speech act to bring it into the format required by a preverbal message.

Sections 4.3 and 4.4 are devoted to macroplanning. The former discusses in greater detail how a speaker selects information instrumental to the communicative goal; the latter is concerned with the speaker's linearization problem, i.e., how a speaker orders for expression complex information involving several messages.

Section 4.5 explores some of the processes underlying microplanning. A speaker will mark referents in a message for their accessibility in such a way as to guide the listener's attention to what is already given in the discourse or to signal that a new entity is being introduced. He may also want to mark a particular referent as the topic. The speaker must further take care that all information is given the necessary propositional format, and that each preverbal message acknowledges the language-specific requirements of the Formulator.

## 4.1 From Intention to Message

The mother of each speech act is a communicative intention. This notion was introduced in chapter 2. The speaker's "proximal" purpose in planning a speech act, we saw, is *intention recognition* by the addressee. The present chapter discusses how a speaker goes from a communicative intention to a preverbal message that, when formulated, will make the cooperative addressee recognize that intention. We will be only marginally concerned with other, more "distal" intentions of speakers, which are not to be recognized as such by the other party. It should, however, be added that such more distal intentions are by no means uninteresting for a theory of the speaker; they may affect the selection of information to be expressed (see subsection 4.3.5) as well as the prosodic features of speech, such as loudness, rate, or intonation.

Let us begin with an example, a case of informing. Speaker Simon wants to tell hearer Hanna that Wubbo is an astronaut. More formal, Simon's message encoding begins with the intention to bring about a situation in which

KNOW (HANNA, INTEND (SIMON, BELIEVE (HANNA, ASTRONAUT (WUBBO)))).

The goal state is Hanna's knowing that Simon intended her to believe that Wubbo is an astronaut. There may be several ways for Simon to achieve this goal state. They need not even be verbal means. Simon could, for instance, show Hanna a picture with Wubbo in an astronaut's outfit. If Hanna knows Wubbo but didn't know that he is an astronaut, Simon's showing the picture can fulfill the communicative intention. But if Simon's communicative intention is *illocutionary* (i.e., if the intention is to use *verbal* means to bring about the goal state), Simon must encode a message. In order to achieve his goal, Simon could encode the message DECL (ASTRONAUT (WUBBO)). When formulated, this message would be

uttered as *Wubbo is an astronaut*. If Hanna is cooperative, she will derive not only the proposition that Wubbo is an astronaut but also that Simon intended her to believe this proposition.

The journey from intention to message will, normally, involve more than a single step. Often, the goal will have to be expanded into subgoals. If the intention is, for example, to give a person a route direction, the speaker must create a whole plan consisting of subgoals ("first direct her to the city center, then inform her about the precise location of the museum") and sub-subgoals ("to go to the city center first take the freeway, then turn right at the second traffic light"), and so forth. The speaker will have to plan and order the various subgoals to be achieved. For each of the subgoals he will have to decide on a speech act to be expressed—that is, on an assertion, a command, a question, or whatever. This involves, for each speech act, the planning of information to be expressed to the interlocutor in order to satisfy the goal. As was extensively discussed in chapter 2, much can be conveyed without being explicitly formulated. The speaker will, normally, count on the capacity of a cooperative interlocutor to *infer* the goal or subgoal from an utterance that expresses only a fraction of the information to be conveyed; speech acts can be indirect. The sum total of these activities will be called *macroplanning* (roughly following Butterworth 1980b). Its output is an ordered sequence of what we will call *speech-act intentions* (sometimes shortened to "speech acts"). These are messages as far as specified for intended mood (declarative, interrogative, imperative) and content.

But the speaker must not only plan and order the contents of successive speech acts. The contents of each speech act should also be brought into perspective, and a particular information structure has to be assigned. The distribution of what should be expressed as topical, focused, or new information must be assigned, and the speaker will have to acknowledge certain language-specific requirements the message has to satisfy. These activities on the part of the speaker will be called *microplanning*. The output of microplanning is, for each intended speech act, a preverbal message.

It is tempting to view macroplanning and microplanning as two stages in the process of message encoding. In macroplanning a speaker elaborates a communicative intention down to the level of the content of individual speech acts; in microplanning the content of each intended speech act is given informational perspective and is assigned all the features that are obligatory for a preverbal message. Such a two-stage theory should, of course, be of the incremental sort. It is not necessary for a speaker to complete all macroplanning before microplanning can start. One can begin

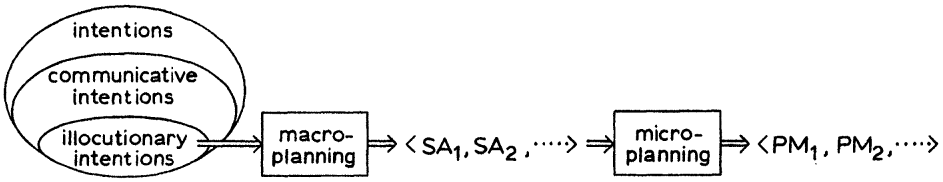



**Figure 4.1**
From intention to preverbal message.

giving a route direction without having planned all its details. This two-stage view may eventually turn out to be too restrictive; there may be situations in which macroplanning is affected by microplanning. Until then, however, the two-stage view is to be preferred because it is more restrictive.

The issues discussed so far are summarized in figure 4.1, which shows that communicative intentions are a subclass of intentions and that only a subset of communicative intentions are to be realized by means of speech acts. There are, then, two kinds of process involved in message encoding: First, the speaker must elaborate the intention. For each of its subgoals, the speaker should plan a speech act (SA)—i.e., should select information whose expression would be instrumental in realizing the goal. This is macroplanning. Second, each informational unit to be expressed must be shaped into a preverbal message (PM) by giving it an information structure, a propositional format, and a perspective that will guide the addressee's attention in the intended way and meet the input requirements of the Formulator. This is microplanning.

Both sides of the message-encoding process are heavily context-dependent. In order to elaborate his goals and to select *effective* information for expression, the speaker must take into account the precise discourse situation, and the same holds for the assignment of informational perspective. A discourse context is a continuously changing situation, and the processes of message encoding must therefore refer to the records the speaker keeps of the ongoing discourse. We will take up this important issue of the speaker's discourse records before we turn to the processes depicted in the figure.

## 4.2 Bookkeeping and Some of Its Consequences for Message Construction

The defining characteristic of coherent discourse is that every new move of a speaker is in some way related to whatever was said before. A cooperative speaker's contributions are supposed to be relevant to the ongoing dis-

course. But this requires bookkeeping or storage on the part of the speaker. He will have to keep track of what was said and what was conveyed by himself and by the interlocutors. The sum total of the information about the discourse that is currently available or accessible to the speaker is called the speaker's *discourse record*; it is the speaker's internal representation of the discourse as it evolved.

A discourse record is by nature a dynamic entity. It changes with each new contribution made to the conversation, whether by the speaker or by another participant. The speaker's record is not just a superficial trace of all utterances made; it is a structured interpretation of what happened in the conversation. An interesting but little-studied issue is which aspects of discourse lead to deep encoding (i.e., to long-term storage) and which aspects are transient (i.e., kept in working memory for only short periods of time). Let us review some of the major titles in the speaker's account books.

### 4.2.1 The Type of Discourse

There are different types of discourse, and they require different kinds of contributions on the part of the speaker. There is, first and foremost, *informal everyday conversation.* This may be a less unified type than is generally supposed. For instance, it is doubtful that everyday chatting with peers is of the same nature as everyday chatting with parents or children. Among the defining characteristics are, at any rate, the interlocutors' awareness of informality, of roughly equal rights to the floor, and of the freedom to change topic. When a speaker experiences the discourse as everyday conversation, he tacitly knows which turn-taking rules to follow—namely, the ones discussed in chapter 2.

Analysts of conversation have distinguished and studied various other types of discourse. *Narrations* (Beaugrande 1980; Labov 1972; Scollon and Scollon 1981)—including the telling of stories (Jefferson 1978; Ryave 1978) or dirty jokes (Sacks 1978)—can happen inside everyday conversations, but they require an awareness on the part of the interlocutors that a single speaker has the preferential right to the floor until the narrative is completed. The speaker can count on the suspension of disruptive self-selection by other participants, but has to pay by having to generate a structured sequence of messages. *Lectures* (Goffman 1981) are similar in the latter respect, but they are usually not embedded in conversation; they are more serious and impersonal. The speaker is supposed to impart his views to a spatially marked audience. *Examinations* and *interviews* (Atkinson and Drew 1979) are characterized by a fixed question-answer turn order, in which roles are clearly divided as to who does the questioning and who

gives the answers. *Debates* (Walton 1982) share much of the question-answer nature of examinations, but the role of questioner is now equally distributed between the participants, and each participant has some thesis to defend. This list is by no means complete. The many other types of discourse include *planning discourse* (Linde and Goguen 1978; Goguen and Linde 1983), *route direction* (Klein 1982; Wunderlich and Reinelt 1982), *spatial description* (Ehrich 1982; Ehrich and Koster 1983; Levelt 1981, 1982a,b; Linde and Labov 1975), *radio talk* (Goffman 1981), and *therapeutic discourse* (Labov and Fanshel 1977). A good source on kinds of discourse is volume 3 of van Dijk 1985.

The speaker has to keep track of the type of discourse in which he is engaged, and of the special role assigned to him. The speaker makes a category mistake if he constructs his messages in the framework of the wrong discourse type—for example, if he takes an examination to be a debate, or air-traffic-control discourse to be everyday conversation. It is especially important that the type of discourse in which interlocutors are engaged be mutually known, so that the participants will be on common ground. Establishing agreement on the discourse type may require explicit negotiation at the outset, but usually the type of discourse is *invoked* by the way the talk is conducted (Schegloff 1987). For instance, it is in the way one person talks like a doctor (i.e., speaking of a "hematoma" instead of a "bruise") that the interlocutor recognizes that the discourse is of the doctor-client type.

### 4.2.2 The Topic of Discourse

The *discourse topic* is what is being talked about, and thus mutually experienced, by the participants. A conversation can be about the cost of living, an interview about one's medical condition, a planning discourse about tonight's burglary, and so on. The maxim of relation requires the speaker to make his contribution relevant to the ongoing discourse. Normally, a contribution will relate to the discourse topic. That is what makes discourse coherent. Brown and Yule (1983) call this "speaking topically."

But a speaker may want to change the topic in order to realize some goal not related to the present topic. This involves establishing agreement with the interlocutors about the topic shift. There are myriad devices with which a speaker can initiate a shift of topic. It can be done by explicit declaration (*The next issue I want to address is . . .*); this is what Grosz and Sidner (1985) call a *true interruption*. It may also be done unobtrusively by suggesting a relation to the current topic (*That reminds me of . . .*); this introduces a *digression*. A *flashback* (*Whoops, I forgot to tell you that . . .*), finally, is a

kind of interruption a speaker will make in order to satisfy some subgoal that should have been handled at an earlier stage; it is usually followed by a return to the main topic. And there are prosodic means of marking the introduction of a new topic, such as raising the pitch (Brown, Currie, and Kenworthy 1980). The choice of means for effecting a shift will also depend on the type of discourse, and especially on the distribution of power among participants in that type of discourse.

Brown and Yule (1983) point out that a speaker's topic may or may not develop into a discourse topic, depending on whether it meets with mutual agreement. The speaker has to keep track of what is the current discourse topic, and he should not confuse it with his private "speaker topics." It is exactly confusion of this kind that makes schizophrenic speech incoherent (Brown 1973; Rochester and Martin 1979).

Though the notion of discourse topic is intuitively obvious and its relevance for discourse understanding has been demonstrated over and again (Bransford and Johnson 1973; Sanford 1985), it is notoriously hard to formalize. When a conversation is about the cost of living, it can at the same time be about the cost of food, and about the cost of bread. Most interesting topics of conversation are hierarchical in structure, and participants in a discourse can move up and down through the hierarchy in making their contributions. Grosz and Sidner (1985) relate this to the intentional structure of the discourse, i.e., the hierarchy of goals and subgoals being developed. When a subgoal is introduced, interlocutors descend a step in the hierarchy; when the subgoal is satisfied, one can return to considering a higher-level goal. Large moves are like topic shifts, and may require some negotiation; small moves occur from utterance to utterance.

The lowest-level subgoals eventually evoke individual messages and their topics (see subsection 3.4.2); they are the smallest ingredients in the hierarchy of discourse topics.

A speaker who keeps track of what is being talked about doesn't do so by keeping a running but unstructured list of topics. What is kept is (i) a structured mental representation of the discourse content as it developed and (ii) a pointer to the part of the content structure and the goal structure that is now being elaborated. Let us consider these two elements in turn.

### 4.2.3 The Content of Discourse: Discourse Models and Presuppositions

Interlocutors introduce and reintroduce referents (persons, things, events, etc.) and make predications about them. In doing so, they build "mental models" (Johnson-Laird 1983) of these entities, their relations, and their

properties. A *discourse model* is a speaker's record of what he believes to be shared knowledge about the content of the discourse as it evolved (Johnson-Laird and Garnham 1980; Kamp 1984; Prince 1981; Seuren 1985; Webber 1981).

Discourse models are populated by representations of token events, persons, and so forth—i.e., entities to which reference can be made. Also, it is possible to make reference to finite or infinite *sets*, as in *the world wars* or *the natural numbers*. Each of these referents is said to have an *address* in the discourse model. And when something is communicated about that referent, the predication is added to its address.

Every new speech act of a participant changes the state of the discourse model. It may add an address, which is typically done by using an indefinite expression. If a speaker says *there is a baby in the bath*, he invites his interlocutor to add BABY as an address but he is not yet *referring* to any particular baby. The predication is true if any baby is in the bath. With respect to the discourse up to that point, this is *new information*.

A speech act may also add information to an existing address. The former speaker's next sentence may be *the baby is crying*. The speaker is now making reference to a particular baby. The sentence is true only if the baby in the bath cries, not just any baby. Typically, a definite expression ("*the* baby") is used for referring to an already existing address. The baby is now *given information*. (See subsection 4.5.1 for an elaboration of these notions.)

Depending on their intentions, speakers may convey information concerning existing or newly introduced entities or may invite other participants to provide such information. Put more formally: Speech acts can be conceived of as mappings of the current discourse model onto the next. They change the shared set of referents, or the shared information about them. The shared information includes, in particular, information about the interlocutors. When a speaker gives a command, for example, he adds to his information about the addressee that the addressee now knows his intention to commit her to a certain action.

The speaker's utterance invites the addressee to infer the communicative intention, i.e., to construct a representation of the information to be conveyed. This may involve the creation of new addresses and the storage of new information under old or new addresses. Normally, the speaker's purpose will be that the listener's representation agree in essential points with his own.

In the simplest case—two-person interaction—there are four knowledge structures involved in this part of a speaker's bookkeeping. To distinguish

among these, let us consider a fictitious example in which Marcia and Seth are conversing about their travel experiences. The topic of discourse is, more specifically, travels in Italy. Marcia's and Seth's mental representations are depicted in figure 4.2. The nodes represent addresses for referents, such as places, persons, and events. The figure represents only some of the major tokens referred to in the conversation. Many are left out (for instance, the egos of the two participants). What is said about the referents is not really represented either. It should consist of lists of predications stored under each address. In the figure, arcs have been drawn between nodes that figure in the same proposition. If Seth says *the food is nice in Rome*, an arc is drawn between the nodes for FOOD and ROME, suggesting that the predication relates them in some way in the discourse model.

The first kind of knowledge that is relevant to a speaker's discourse planning is the knowledge the speaker believes he shares with the listener, independent of the present discourse interaction. Let us call this *common ground*. For example, Seth may believe that Marcia knows about the existence of the pope and about the fact that the pope lives in Rome. Figure 4.2 represents for both Seth and Marcia the entities POPE and ROME, as well as the arc connecting them (which stands for their locative relation). POPE and ROME are inside the balloon labeled "common ground." The common ground may also include knowledge of a less general type. Seth and Marcia may have shared knowledge about a common friend, Harry. Seth may, moreover, believe that they mutually know that Marcia wants to talk about Italy, and so on. It is irrelevant for the present discussion whether the common ground is *really* shared by the interlocutors; what matters is that the speaker believes it is. A last item in the common ground is the shared context of discourse—the local scene the speaker believes to be sharing with the interlocutor. Seth knows, from perceptual evidence, where Marcia is located in the scene and how she is oriented. He believes that Marcia shares this knowledge, and that she knows that Seth knows it, and so forth.

The second knowledge structure is what the speaker believes to have successfully conveyed to the listener during the discourse up to now; it is the shared knowledge arising from the speaker's *own contributions*. Seth, for instance, talked about his experiences in Rome but not about the pope. Rather, he discussed the food, the tipping habits, and other matters that he believed Marcia didn't know about. This information, which Seth believes he has conveyed to Marcia, is represented in figure 4.2 as own contributions in Seth's representation. Marcia's own contribution concerns information she conveyed about Florence (where their common friend Harry lives) and

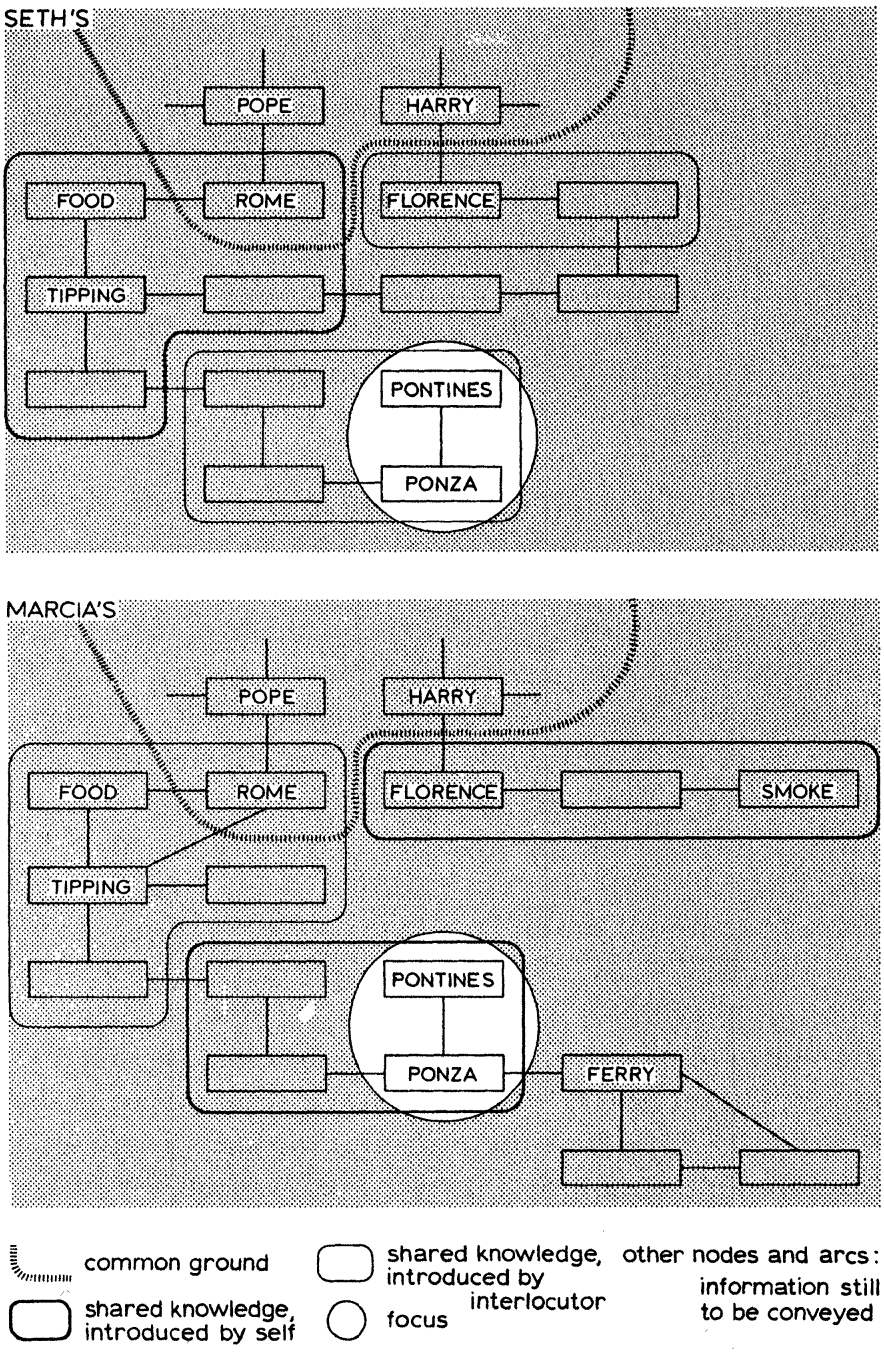



**Figure 4.2**
Types of knowledge involved in example conversation.

about Ponza (an island previously unknown to Seth). These items are demarcated as own contribution in Marcia's representation.

The third knowledge structure is the information the speaker believes the interlocutor to have intended to convey to him in the discourse as it developed. From the speaker's point of view, this is shared knowledge emanating from the interlocutor; it is the *interlocutor's contribution.* Seth, for instance, takes it that Marcia talked about Ponza (an island he had never heard about) and Florence (where their common friend Harry lives). Ideally, the own-contribution part of a speaker's mental representation should be identical to the interlocutor's part of the addressee's mental representation. However, a speaker's belief that this knowledge is shared may be mistaken, and thus may lead to later misunderstanding. Marcia, for instance, believes that she mentioned the smoke poisoning the air in Florence; however, Seth failed to register this, owing to his inattentiveness.

The fourth knowledge structure is information which the speaker still intends to convey but which has not been up for expression yet. This *information to be conveyed* by the speaker is also depicted in figure 4.2. Marcia wants to say more about Ponza's ferryboats, while Seth wants to relate what Marcia said about Florence to his own experiences in Rome. The knowledge intended to be conveyed—i.e., the communicative goal—develops by plan or association as the discourse proceeds.

The speaker's discourse model can now be defined as "own contributions plus interlocutor's contributions"—that is, the knowledge structure the speaker believes he has conveyed to the interlocutor plus the knowledge structure the speaker believes the interlocutor intended to convey in the discourse up to now. The picture will obviously be more complicated for multiparty discourse, but that will not concern us here.

Every move in the conversation changes the knowledge pattern of the interlocutors. When Seth introduces a new entity and believes that Marcia has grasped it, Seth's "own contribution" part is expanded by that element. If indeed Marcia has grasped it, her "interlocutor's contribution" part is correspondingly expanded, and so on. Discourse models change continually.

It is characteristic of coherent discourse that a new contribution relates to what was said before, i.e., is relevant to the current state of the discourse model. It will either add further linking information to existing addresses or introduce new referents by linking them to existing ones. An important notion here is *presupposition.* To maintain coherent discourse, the presuppositions of a new message must be satisfied or satisfiable in the discourse model. If Marcia were to say *The ferry is convenient* without

having previously referred to a ferry, her contribution would violate a so-called *existential* presupposition, because her use of a definite expression presupposes the existence of a FERRY address in the discourse model or in the common ground. If Seth were to ask *When did you stop taking your vacations in Italy?*, there would be the *factive* presupposition that Marcia had indeed stopped taking her vacations in Italy. Such a question is a coherent contribution only when that fact is somewhere in the discourse model or in the common ground.

Still, speakers often ignore presuppositions in making their contributions. They rely on the interlocutor's intelligence to realize what Seuren (1985) calls *backward suppletion* of the presupposed information. When Marcia talks about *the ferry*, Seth will derive that there must be some ferry, and he will set up a provisional address for it in the hope and expectation that he will soon be informed about which ferry Marcia means. Similarly, Marcia, in response to Seth's question, will think "Ah, he thinks I am not going to Italy anymore". Marcia adds Seth's apparent belief to the discourse model, and may then either deny it when it is false (*Oh, I haven't stopped going*) or answer the question if it is a true belief (*Since 1985*).

But backward suppletion is impossible when information contrary to the presupposition is already in the discourse model. In that case the presupposition is really violated, and the discourse is incoherent. This happens in the following bit of discourse.

Marcia: I always go to Italy for vacation.

Seth: When did you stop taking your vacations in Italy?

This led Seuren (1985), following van der Sandt (1982), to propose the following presupposition test: If S is the formulation of a message, then P is the formulation of a presupposition of the message if and only if

(i) P, but/and S is coherent

and

(ii) perhaps not P, but/and S is incoherent.

The presupposition of Seth's question can be formulated as *You stopped taking your vacations in Italy*. The first test requires the following sequence to be coherent: *You stopped taking your vacations in Italy, and when did you stop doing so?* It clearly is coherent. The second test predicts incoherence for the sequence *Perhaps you didn't stop taking your vacations in Italy, but when did you stop doing so?*, which is incoherent indeed. Therefore, Marcia's having stopped taking her vacations in Italy is indeed a presupposition for the message in Seth's question. The same pair of tests can easily be applied

to the earlier existential presupposition about the ferryboat. In generating a message, a cooperative speaker will introduce referents and will maintain reference in such a way that no violation of presupposition results.

It should be obvious that, for any sizable discourse, the speaker cannot keep the whole discourse model in his center of attention. The discourse model is, rather, stored in long-term memory. But it is easily accessible. At any one moment, only a small fragment of this information is attended or "pointed" to. This fragment is called the *focus*.

### 4.2.4 The Focus

The attention span of a speaker is fairly limited. He can work on only one or a few notions at a time in planning his discourse. And the same holds for the listener who is interpreting the speaker's utterances. The information to which the speaker is attending at a particular moment in time is called his *focus*. Similarly, there is a listener's focus. If the speaker believes that their foci are shared, there is a unique part of the discourse model which the speaker marks as "jointly attended to." In the conversation between Seth and Marcia, this can be the food in Rome, the island of Ponza, or any of the other discourse entities. There is, of course, more in focus than just individual referents. The speaker focuses on a particular goal or subgoal to be satisfied; there is some communicative intention to be jointly attended to.

The notion of attentional focusing is classical in psychology, and is not limited to the study of language use. Wilhelm Wundt (1896) called it apperception; William James (1892) called it apprehension or primary memory. These and many other authors have argued that one can consciously attend to only a few disconnected elements at a time. Miller (1956) spoke of a "magical number seven plus or minus two," and Broadbent (1975) of a "three-slot register." But when information is structured, the span of attention can be larger—and that is often the case when the speaker is planning a message. Our present use of "focus" in the planning of discourse agrees with the way in which Chafe (1979, 1980), Grosz (1981), Grosz and Sidner (1985), Herrmann (1983), and Sanford and Garrod (1981) use the term "focusing."

The encircled parts of the networks in figure 4.2 are supposed to be "currently in focus" for the interlocutors. Seth and Marcia are in the happy circumstance that their foci are aligned. They are both concentrating on Ponza and the Pontines and the information being conveyed about them.

One can further distinguish a *focal center*, as proposed by Grosz, Joshi, and Weinstein (1983). This is what the speaker is newly attending to in making the current utterance; it is the most salient part of his focus. It is

more the rule than the exception that the speaker's focal center differs from the hearer's. By his utterance the speaker is trying to make the hearer attend to something she is not yet attending to. It is, therefore, important to distinguish between the *speaker's focal center* and the *hearer's focal center.* It is a major concern for the speaker to align the focal centers. When the speaker shifts his focal center to a new referent, his utterance should tell the interlocutor where to go; the speaker's and the interlocutor's foci are not yet aligned. In the example, Marcia may shift her focus to the Ponza ferry, intending to say something about it. If she believes that Seth's focus is still as depicted in figure 4.2, she may say *The island can be reached by an old ferryboat.* In this way Marcia connects the new information to what Seth is presently focusing on (Ponza island), and by using an indefinite expression (*an old ferryboat*) she signals that the newly focused element is not yet in the discourse model or in the common ground.

More generally, the speaker's way of instructing the listener where to focus next is dependent not only on the speaker's new focal center but also on the listener's current focal center. The cooperative speaker thus has to keep track of the interlocutor's current focus and its current center. They are represented in the discourse model. Whether the representation is veridical is not relevant. "The interlocutor's current focus" and "the interlocutor's focal center" are what the speaker *believes* to be the interlocutor's focus and focal center—that is, what the speaker will take into account in constructing messages. Whether his belief is warranted is not at issue.

A newly focused element will usually not be the topic of the speaker's message. In *The island can be reached by an old ferryboat,* the message is about the island Ponza, whereas the speaker's new focal center is the ferry. Still, the topic and the new focal center may coincide. Seth could, out of the blue, say *The pope didn't say much in five languages.* The pope is newly focused here, and he is the topic of the message.

Usually, speakers place newly focused or newly introduced information later in the sentence than information that is currently focused or is already in the discourse model. Clark and Haviland (1977) have shown how this principle of ordering helps the addressee to connect new information to tokens currently or recently shared between speaker and hearer. The Formulator will, of course, have to "know" all this. It will, in other words, be necessary for the speaker to encode in the message where the referents can be found—i.e., in the current focus, elsewhere in the discourse model, or in the common ground—or that they are "brand new." We will turn to this issue in subsection 4.5.1.

### 4.2.5 What Was Literally Said

Interlocutors listen, in the first place, for content. But there is evidence that, within limits, they also register the literal wording of what was said. There has been substantial laboratory research on listeners' memory for spoken text (see Levelt and Kempen 1975 and Levelt 1978 for reviews). The major finding is that it is short-lived. Recall and recognition are generally very good for the last clause or sentence heard, but quickly decline to nearly chance level for less recent materials (Jarvella and Herman 1972).

However, three studies performed in more natural settings and with more naturalistic materials have demonstrated that the literal wording of conversational materials affects long-term memory. Kintsch and Bates (1977) tested students' memories of statements from a lecture they had attended and found especially good memory for the literal form of jokes and for statements extraneous to the topic of the lecture. These effects were still measurable five days after the lecture. Keenan, MacWhinney, and Mayhew (1977) recorded the conversation of a lunchroom discussion group and tested memory for the literal wording of what they called "high interactional content" statements. (This notion is somewhat vague; what it means is that there is a high personal involvement of speaker or interlocutors in a statement such as *Do you always put your foot in your mouth?*.) Literal memory was much better for such statements than for "low interactional" statements. These two studies show that listeners do register the literal form of utterances that are interactionally salient. Bates, Masling, and Kintsch (1978) used conversations from a television drama as test materials and specifically analyzed literal memory for the ways in which referents were introduced or maintained in the video conversations. Among many other things, they found that their subjects had better memory for statements in which a person's name was used than for statements in which a personal pronoun was used, and that elliptical sentences were remembered less well than nonelliptical ones. These results seem to show that when the pragmatic role of an utterance is to introduce new referents (which is not done by way of pronouns) or to introduce new statements about referents (which is not done by elliptical clauses), there is more than chance memory for the literal wording used.

Surprisingly, there has been almost no research on what speakers remember of what they have said themselves. Deutsch and Jarvella (1983) compared memory for own speech and interlocutor's speech and found that, under certain restricted experimental conditions, the recall for self-produced speech is better than that for other-produced speech. But this difference can be explained entirely by the speaker's giving more atten-

tion to the *content* of self-produced speech than to the *content* of other-produced speech. Or, in terms of discourse models, a speaker will keep a better record of self-conveyed information than of interlocutor-conveyed information. The fact that one's own speech is *self-articulated* may be entirely irrelevant.

Does memory for what was literally said have any significance for what a speaker is going to say? There is observational (Schenkein 1980; Harley 1984) and experimental (Levelt and Kelter 1982) evidence that the wording of the interlocutor's last turn can affect the wording of the current speaker's turn. In one of Levelt and Kelter's experiments, shopkeepers were telephoned and asked the Dutch equivalent of one of the following four questions:

(1) What time do you close?

(2) At what time do you close?

(3) What time do you close, since I will have to come downtown especially for this, you know?

(4) At what time do you close, since I will have to come downtown especially for this, you know?

These questions differ in two ways: They contain or do not contain the preposition *at*, and they are short (without additional clause) or long. The preposition *at* (or more precisely the Dutch equivalent thereof) has no meaningful function whatsoever. The answers were scored for the presence of *at* in the shopkeeper's time phrase (e.g., *five o'clock* or *at five o'clock*). It was possible for *at* to occur or not to occur in an answer to any of the four questions. For questions 1 and 2, it turned out that the answers tended to follow the question in the use of *at*. Question 2 elicited significantly more *at* answers than question 1; question-to-answer correspondence was 61 percent. In other words, the interlocutor's *at* must have been registered by the shopkeeper, in spite of the fact that it had nothing to do with the content of the question. For questions 3 and 4, however, there was no systematic question-to-answer correspondence at all; it dropped to a random 47 percent. This makes it likely that, in conversation, literal recall not supported by salient content or pragmatic significance is short-lived, probably going back only as far as the last clause uttered.

When, however, the literal wording *is* important for conversational purposes, speakers will make a record of it and use that wording with profit. Clark and Wilkes-Gibbs (1986) had subjects communicate information about a set of tangram figures—irregular geometrical shapes that are hard to name. After some give and take, interlocutors settled upon a fixed

referring expression for each of the shapes (e.g., *the angel*). This was conversationally far more practical than giving again and again the full initial descriptions of the same figure (e.g., *sort of an angel flying away or something; it's got two arms*). Of course, this requires the speaker to maintain a record of the literal expression that was used previously by himself or by the interlocutor.

This completes our list of the items in the speaker's discourse record. It will become apparent from the following sections that the speaker refers to these records continually, both in macroplanning and in microplanning.

## 4.3 Macroplanning 1: Deciding on Information to Be Expressed

Illocutionary intentions are intentions to commit oneself or an interlocutor to the factuality or desirability of something. The speaker's utterances should make these intentions recognizable to the person concerned. Depending on the character of the illocutionary intention, this may involve the construction of one or more speech acts. When the intention is to inform an addressee that a train is arriving, a single speech act may suffice (namely the assertion that the train is arriving). When, however, the intention is to inform the hearer about the route from Florence to Rome, a sequence of several declarative speech acts may be necessary. Conversely, one speech act can realize several communicative intentions at the same time. When Marcia asks Seth whether he met Harry recently, Seth may answer *I saw the bastard in Florence*. This answer communicates to Marcia both a commitment to the factuality of Seth's having seen Harry in Florence and a commitment to an opinion about the character of Harry. This many-to-many mapping from communicative intentions onto speech acts complicates the analysis of macroplanning greatly.

The major requirement in order for the information to be expressed in the speaker's message is that it be *instrumental* in changing the addressee's discourse model in the intended way. An instrumental message need not express each and every detail of the information to be conveyed to the listener. The speaker will assume that the cooperative listener will be able to *infer* the communicative intentions from well-chosen bits of information expressed in the message. The relations between the information to be conveyed and the information to be expressed are governed by Gricean principles (discussed extensively in chapter 2). In particular, the maxims of quantity require the speaker to sail a middle course between being overinformative and being underinformative. Information that is readily inferable from shared knowledge—be it in the discourse model or in the

common ground—should, as a rule, not be expressed in the message; the speaker will transmit it by implicature. When, in a route description, a speaker says *Then turn right at the church*, he conveys by implication that there is some road or pathway there. This need not be expressed; the intelligent addressee can infer it. It is, indeed, instrumental to convey information by implication. First, it is efficient; the speaker will have to express only a part of the whole package of information. Herrmann (1983) calls this the *pars-pro-toto principle*. Second, it is a way to acknowledge the addressee's intelligence and cooperativeness. Third, it can be a way to express secondary communicative intentions, such as the intention to be polite. Finally, expressing every detail of the information to be conveyed would induce the interlocutor to interpret the speaker's contribution as flouting a maxim of quantity, and this may lead to inferences not intended by the speaker.

In the following, we will first consider the format of macroplanning procedures. We will then turn to empirical psycholinguistic studies of macroplanning—in particular, the use of attentional resources in elaborating communicative intentions, the speaker's selection of information for making reference to objects, and the construction of requests.

### 4.3.1 The Format of Macroprocedural Knowledge

Chapter 1 described procedural knowledge as a collection of condition/action pairs of the following form: IF a certain set of conditions are met, THEN perform a certain action. What sort of IF/THEN pairs are involved in macroplanning? In other words, if macroplanning is procedural in nature, what form will these procedures have? In view of the assumed process depicted in figure 4.1, the condition for a procedure should, at least, involve one or more illocutionary intentions. And the action to be performed must be a speech act, or a string of speech acts (such as assertions or questions). Only students of artificial intelligence have begun to formulate such procedures explicitly (see especially Appelt 1985), and they do not claim psychological reality for their proposals. Still, it is worthwhile to consider one or two such theoretical proposals before we turn to empirical studies of macroplanning.

Let us first take up the simplest cases of informing. Section 4.1 gave the example of a speaker, Simon, who had the illocutionary intention to inform a hearer, Hanna, that Wubbo is an astronaut. Simon realized the intention by asserting that Wubbo is an astronaut. This can be put in more general terms (roughly following Appelt 1985): Assume that the speaker's illocutionary intention is to bring about the state

KNOW(H, INTEND(S, BELIEVE(H, P))),

where H is the hearer, S the speaker, and P the proposition to be believed. One way of realizing this intention is to assert P. This, in turn, can be done by encoding the message DECL(P). In fact, this is a general message-encoding procedure for informing:

IF the goal state is KNOW(H, INTEND(S, BELIEVE(H, P))), THEN encode message DECL(P).

In other words, it is part of the speaker's procedural knowledge that the speech act of asserting P will, under certain conditions, bring about the effect that the hearer believes that the speaker intended her to believe that P is the case (for instance, that Wubbo is an astronaut). The conditions have to be further specified. The speaker must, among other things, believe that the hearer can hear him. The speaker must, moreover, believe that the hearer doesn't yet know P. Simon will probably not assert to Hanna that Wubbo is an astronaut when this is already mutual knowledge. If he does, he is probably realizing another communicative intention, for instance to reveal to Hanna that he envies Wubbo. These conditions, and probably several more, must be added to the IF-statement above.

A similar condition/action pair can be formulated for the intention of requesting some action. Simon may want Hanna to buy a stamp. The intended goal state is that Hanna know that Simon intends her to intend to buy a stamp. A speech act realizing this intention is a request. More formally:

IF the goal state is KNOW(H, INTEND(S, INTEND(H, DO(A)))),
THEN encode message ?(FUTURE (DO(H, A)))),

where A is the intended action.

If this is part of Simon's procedural knowledge, he will ask Hanna *Will you buy a stamp?* But here, too, there are additional conditions. There is the same physical condition as before: It must be mutual knowledge that Simon believes that Hanna can hear him. Also, Simon must believe that Hanna doesn't yet intend to buy a stamp. If Hanna's intention to buy a stamp is already mutual knowledge, requesting that Hanna buy a stamp involves another communicative intention (for instance, to let her know that she should hurry a bit). In other words, several IF-conditions will have to be added for this procedure to work. In fact, the conditions for different request forms (*Can you buy a stamp?*; *Please, buy a stamp*; etc.) are quite complex, as is apparent from psycholinguistic studies on requesting (see subsection 4.3.3). There is, obviously, still a long way to go in the explicit

formulation of even the most elementary macroplanning procedures. Let us now turn to some empirical studies of macroplanning.

### 4.3.2 Macroplanning and Attentional Resources

Which information has to be conveyed to the addressee depends entirely on the speaker's intentions and on the current state of the discourse model. The information may already be in the speaker's focus. It may, alternatively, involve an elaborate memory search. An example of the former is a situation in which a speaker happens to look out a window, notices a rainstorm, and develops the intention to inform an addressee about this. The event is in the speaker's focus; no further information has to be retrieved. An instance of the latter is a situation in which another party asks for a route direction. A cooperative and knowledgeable speaker will aim at filling the gap in the interlocutor's discourse model by retrieving the locative information that constitutes the path from "here" to the goal place. This may involve several steps of retrieval. The speaker will use the source and goal positions as retrieval cues for accessing a relevant part of his cognitive map of the town, which is in his long-term memory. He will then infer a shortest or easiest route connecting the two positions. He will retrieve landmarks, such as churches and viaducts, to identify successive parts of the route for the addressee, and so on. Retrieving information to be expressed can involve substantial memory search, inference, and planning.

Elaborate search is characteristic for certain discourse types. Not only route directions, but also narrations, lectures, speeches, and other monological forms of discourse are in this category. Interviews and debates are other examples. Such planning and search is under executive control and requires the speaker's attention. At the same time, the speaker must keep some attention available for the further preparation of each message—the microplanning—so that speech can keep flowing while information is being retrieved.

There is some evidence that in longer monologues speakers slowly alternate between phases in which they spend much attention on information retrieval and inference (i.e., on macroplanning) and phases in which they concentrate on finalizing messages for expression (i.e., on microplanning). This evidence stems from work by Henderson. Goldman-Eisler, and Skarbek (1966), Goldman-Eisler (1968), Butterworth (1975, 1980b), and Beattie (1983) on the distribution of pausing and speaking in longer stretches of speech or monologues. (See Petrie 1988 for a different kind of experimental evidence.)

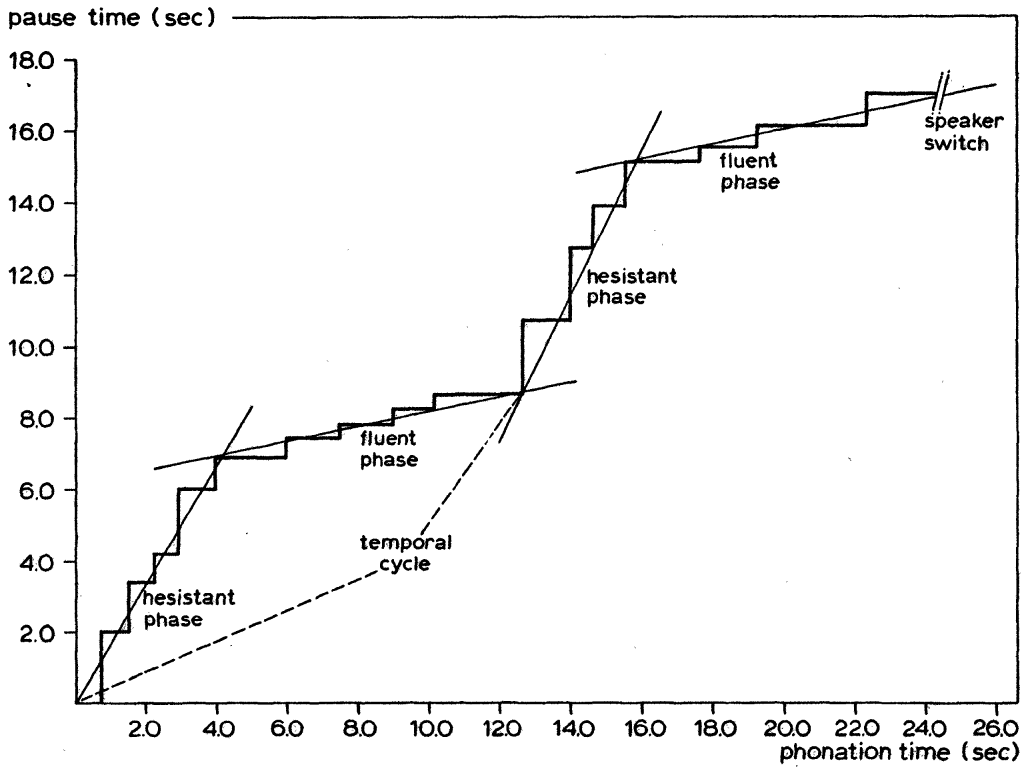



**Figure 4.3**
Alternation of hesitant and fluent phases in a monologue. (After Beattie 1983.)

Henderson et al. (1966) represented speech-pause alternations in what I will call "Henderson graphs." One such graph is presented here as figure 4.3. Both coordinates represent time—the horizontal axis speaking time, the vertical axis pausing time. Each successive segment of either speech or silence is traced in the corresponding horizontal or vertical direction. What counts as silence is any nonphonation period longer than 200 or 250 milliseconds (200 in the case of figure 4.3). The general slope of the resulting curve is the ratio of silence to speech for the stretch of talk represented.

The relevant observation by the above-mentioned authors is that there seems to be a rhythmic alternation of slope in Henderson graphs; steep and flat phases follow one another in a rather regular fashion. The steep parts are called phases of *hesitant speech,* since there is much pausing. These authors also suggest that this pausing is due to speaker's attentional preoccupation with goal elaboration and information retrieval—with macroplanning. The flat parts represent phases of *fluent speech,* with relatively little pausing. Jaffe, Breskin, and Gerstman (1972) and Power (1983, 1984) have warned that these fluctuations may be random, but other studies have made that interpretation less likely (see Beattie 1980, 1983, 1984). The curve in figure

4.3 is derived from a sample of videotaped teacher-student interactions. Beattie (1980) analyzed all single-speaker stretches of talk longer than 30 seconds. These monologues showed a mean cycle time (hesitant phase plus fluent phase) of 22 seconds, with a rather large standard deviation of 16 seconds. Beattie (1980) demonstrated the nonrandomness of these cycles by means of a judgment experiment in which subjects received transcripts of successive sentences from these monologue parts in random order and the task was to reestablish their original order. This turned out to be significantly easier for sentences occurring within the same cycle than for sentences that succeeded one another over a cycle break. In other words, there is more conceptual coherence within a cycle than between cycles. The cycle, it seems, involves the elaboration of some communicative goal into a series of speech acts. This results in a coherent discourse segment. When the speaker then shifts his focus to a new goal or subgoal, a new cycle of elaboration will start, resulting in the next coherent discourse segment.

Though these are suggestive results, one should be careful not to generalize the cycle notion to each and every kind of monologue. If the speaker's attention fluctuates between macroplanning and microplanning, this will lead to overt alternations in fluency only when macroplanning is effortful. Indeed, there have been independent experimental demonstrations of the relation between speech fluency and the "cognitive load" imposed by selecting information for expression. Goldman-Eisler (1968), for instance, showed that there is more fluency in the execution of a cartoon-description task than in that of a cartoon-interpretation task. Good and Butterworth (1980) asked subjects to give route descriptions. One comparison the authors made was between a familiar route (that from home to work) and a relatively unfamiliar one. The familiar route was described slightly but significantly more fluently than the unfamiliar route, as measured by the total percentage of silence time in a description (33 percent and 37 percent, respectively). When a subject was asked to repeat the description of the familiar route, the silence ratio dropped to 27 percent. Selecting information for expression is presumably much easier in the latter case because it is still highly available in memory.

The latter situation approaches one that Clark and Clark (1977) called "ideal delivery"; the subject already knows what he wants to say, and utters it fluently. Clark and Clark suggested that speakers strive for smooth delivery of clauses; they try to minimize within-clause pauses. Good and Butterworth (1980) indeed found that within-clause pauses (over 250 milliseconds long) were substantially less frequent in the repeated descriptions than in the original descriptions (28 percent and 41 percent of the total

number of pauses, respectively). In other words, when the information-selection task is less demanding, the speaker can spend more attention on delivering "clause-ready" messages—i.e., on microplanning.

Attentional lapses in macroplanning sometimes lead to all-out speech errors. Harley (1984) reported several such cases. In one of them the speaker wavered between saying that he got up at 8:52 and that he felt fine at 8:52, and this led to *I felt up fine at 8:52.* Freudian speech errors are also said to be due to attentional lapses, and we will return to them in chapter 6.

### 4.3.3 Selecting Information for Making Reference to Objects

Olson (1970) made the almost trivial but highly seminal observation that a speaker's referring expression indicating some object in the environment is a function of what alternative objects there are in the context of reference. The same object may be referred to differently, depending on the set of contextual alternatives. If a speaker wants to make reference to a big black ball in a situation where speaker and addressee jointly observe only two black balls, a big one and a small one, the referring expression is likely to be *the big one* or *the big ball.* If, however, the alternative object is a big white ball, the speaker's referring expression will be *the black one* or *the black ball.* The choice of attribute depends on whether it can be distinctive for the addressee. A speaker who wants to convey the identity of an object to an addressee will almost never have to list all the object's properties; he will select for expression some subset which is *instrumental* for the addressee to uniquely identify the object in the context of reference.

As was mentioned above, Grice's quantity maxims require the speaker to sail between the Scylla of underdetermination or ambiguity and the Charybdis of overspecification or redundancy. In the above-mentioned situation, an example of underdetermination would be the use of the referring expression *the ball,* which does not discriminate sufficiently; an example of overdetermination would be the use of *the big black ball,* which contains more information than is necessary. What do speakers do in actuality?

The general finding is clear (see Deutsch and Pechmann 1982 for a review): Referring expressions are almost always sufficient or nonambiguous, but they tend to be redundant. The latter seems to contradict a Gricean maxim and needs further scrutiny. An example can be found in the cited paper by Deutsch and Pechmann. They asked children and adults to select from an array of eight toys the one they would most like to give as a birthday present to a child. The toys in the array were of two to four

different kinds (for instance, combs, spoons, cars, and cups); some were large and some small, and they were of three or four different colors. The combinations of these attributes were varied over different arrays. Only 6 percent of the adults' referential expressions were underspecifications, but 28 percent were redundant. In a similar experiment, Pechmann (1984) found that more than 60 percent of adults' referential expressions were redundant.

Pechmann analyzed some of the reasons for this seeming deviancy from the maxim of quantity. In his 1984 study the adult subjects were presented with a sequence of slides, each containing a variable number of objects, which could differ in kind, color, or size. One of the objects in a slide was marked with a star. The subjects' task was to tell an imaginary listener, who would see the same slide but not the star, which object was marked.

Figure 4.4 gives an example of an initially presented array of four objects (Pechmann used blue and red rather than black and white) and three potential subsequent arrays of three objects. For each array, the Dutch subjects' characteristic response is given in English translation; capitalization indicates prosodic accent. The response to array A is typically *the black bIrd.* Neither *black* nor *bIrd* alone would have been sufficient for the

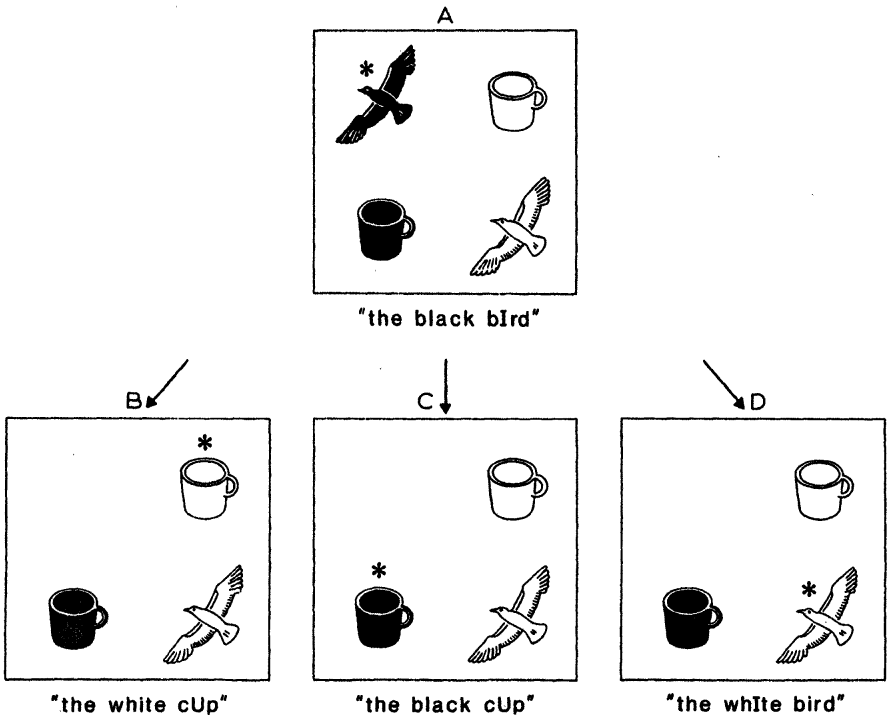



**Figure 4.4**
Stimuli used in an object-naming experiment. (After Pechmann 1984).

listener to identify the object. The referential expression is nonredundant. If it is array B that follows array A, subjects typically respond with *the white cUp*. In this case too both the color and the object name are necessary to distinguish the referent from the alternatives; there is another white object, and there is also another cup. But what happens if either the color or the object name suffices to discriminate the object? In array C it is the color name that would suffice; there is only one black object in the array, and it is the referent. Still, most subjects refer to it as *the black cUp*, in spite of the fact that *cUp* is redundant and moreover nondiscriminating (there is another cup in the array). Also, when the object name is the only discriminating information, subjects tend to include the color name as well. This is the case for array D, where *bird* would be enough. Subjects typically say *the whIte bird* in this case. Note also that for both array C and array D the *non*discriminating information is given prosodic prominence: *cUp* in array C and *whIte* in array D. We will return to this finding shortly.

It would be wrong to conclude from this example that speakers talk redundantly because they simply supply the listener with *all* information concerning the referent. This doesn't hold in general for Pechmann's results, and it is also contradicted by the findings reported in Herrmann and Deutsch 1976, Deutsch and Pechmann 1982, Herrmann 1983, and several other papers. Especially when there are more than two attributes involved, speakers start economizing on their referential expressions. It is also unlikely that speakers introduce some redundancy in their object names in order to help the addressee cope with "degraded communication"; if one discriminating attribute is missed by the listener, he will still be able to identify the referent by means of another discriminating attribute. We saw, however, that speakers also make their descriptions redundant by adding nondiscriminating attributes.

There are at least two other reasons for a speaker to overspecify the referent. One is that redundant nondiscriminating information can help the addressee find the referent. Deutsch (1976), Mangold (1986), and Sonnenschein (1982, 1984) showed that it is easier for listeners to identify an overspecified referent than a minimally specified object in an array of objects. Listeners apparently create a "gestalt" of the object for which they have to search. It is harder to search for "something red" than for "a big red bird", even if the color would be sufficiently discriminating. Information about the *kind* of object to be looked for (e.g., a bird) is especially helpful for constructing such a gestalt. This would explain a general tendency in the experimental findings on object naming to include the noun in the refer-

ential expression even when redundant (e.g., *the red bird* instead of *the red one*). The nature of the "distractor" objects is irrelevant here.

Another reason for the apparent redundancy has to do with the speaker's discourse model. A speaker cannot refer only by mentioning what discriminates the object in the situation perceptually shared with the addressee (i.e., in what was earlier called "common ground"); he can also refer by contrasting it to the last one focused by the listener. In other words, the new focal center can be introduced by expressing information that makes it contrast to what is in the addressee's current focal center. This is presumably what happened for arrays C and D in figure 4.4. We observed that in both cases speakers gave nondiscriminating information (*cUp* and *whIte*, respectively). The speakers apparently contrasted the new referent object with the previous one (*black bird*), in spite of the fact that the old referent was no longer *perceptually* present. That comparison required mentioning the differing feature, whether it was perceptually redundant or not. In other words, in terms of the discourse model *cUp* and *whIte* express the *non*redundant information, whereas *black* and *bird* are redundant because the previous object in the focal center was a black bird. Pechmann (1984) calls this latter kind of redundancy (i.e., with respect to the discourse model) *endophoric* redundancy, as opposed to the *exophoric* redundancy which derives from the set of perceptually given alternatives. Especially noteworthy is that the endophorically discriminating information (*cUp* and *whIte*, respectively) was given prosodic stress. This is a quite general finding in the studies by Pechmann and in those by Terken (1984): There is systematic accentuation of endophorically discriminating information (such as *whIte* in figure 4.4D), but no systematic accentuation of exophorically discriminating information (such as *bird* in figure 4.4D). I will return to this in discussing microplanning in subsection 4.5.1.

What information the speaker will express for referring to one object among alternatives will depend on both endophoric and exophoric factors. Endophoric reference—i.e., reference from the perspective of the previously mentioned object—is especially cooperative when the current referent differs only slightly from the previous one. By stressing the differing feature, the speaker instructs the addressee to just adapt the previous gestalt correspondingly: "Don't construct a totally new template for your search; the previous one, slightly modified, will do." If there are too many differences, however, it is more economical for the addressee to compute a new target gestalt. This is probably the case for the referent in figure 4.4B. It is not efficient for the addressee to construct the target concept WHITE

CUP by deriving it from BLACK BIRD. In that case, exophoric reference is indicated.

Grice's maxims are too general to be experimentally refutable, but the results discussed so far certainly show that there is more than one way to be cooperative in referring to objects. What violates the maxims of quantity from the exophoric point of view doesn't do so from the endophoric. There is still another perspective, which has received surprisingly little attention. Deutsch and Pechmann (1982) pointed out that speakers, especially younger children, underspecify because they *exploit* cooperativeness. When an array contains many similar objects, it would be too much work for a speaker to determine all the relevant exophoric contrasts. Instead he mentions a few salient features, and counts on the cooperative addressee to ask further if the referent cannot be uniquely identified. Deutsch and Pechmann showed that children (and sometimes adults) do this even when they are fully able to distinguish the exophorically relevant features of the referent object.

Clark and Wilkes-Gibbs (1986) turned this vice into a virtue. They stressed that referring to objects in conversation is always a *collaborative* process. Speakers try to establish the mutual belief with their interlocutors that the object reference is understood well enough for the current purposes. And establishing this mutual belief often requires some turn-taking. In the experiment of Clark and Wilkes-Gibbs (already mentioned in subsection 4.2.5), pairs of subjects conversed about arranging irregular shapes for which there are no easy names. Here is a typical effort to establish reference to such a tangram figure:

A: Uh, person putting a shoe on.

B: Putting a shoe on?

A: Uh huh. Facing left. Looks like he's sitting down.

B: Okay.

The first person, A, proposes a referential expression. His partner, B, indicates that it doesn't suffice. A adds some information, and B says *Okay*, creating the belief that he has identified the intended referent. Referring situations are not always as complicated as this one, but Clark and Wilkes-Gibbs argued that simpler cases of referring are not essentially different. The speaker presents a referring expression, and the addressee indicates in some way that it suffices for the current purposes. Isaacs and Clark (1987) provided additional experimental evidence for this theory.

Herrmann (1983) reviewed several more findings about a speaker's selection of information for expression in object reference, such as the finding by Herrmann and Deutsch (1976) that if there are two features,

each of which is sufficiently discriminating in the exophoric sense, speakers tend to choose the feature that captures the most perceptually salient difference between the referent object and the alternatives. The choice, in other words, is perceptually driven. These authors also found strong set effects. If a feature, such as color, has been instrumental time and again, speakers will keep using it even when it becomes redundant. A speaker's cooperativeness displays a certain inertia, apparently.

What is still to be developed is a theory about indirect means of object reference. Chapter 2 mentioned Nunberg's (1979) notion of *referring function*, which relates the demonstratum to the referent. The referent, we saw, need not be the object pointed to. It also need not be the object named. One of Nunberg's examples involved a restaurant waiter going off duty and reminding his replacement that *the ham sandwich is sitting at table 20*. The object named (the demonstratum) is the ham sandwich; the entity referred to is a customer. The demonstratum and the context make it, presumably, possible for the addressee to derive the function and thus the referent. But how does a speaker, who has the referent in mind, discover a referring function that allows him to refer indirectly? And why does he want to do so? What a speaker selects for expression can be quite distinct from the information to be conveyed and can still be instrumental.

### 4.3.4 Selecting Information for Construction of Requests

A speaker who makes a request conveys a multitude of information to the addressee. As was discussed in chapter 1, the addressee should know that the speaker wants her to perform a certain action. But that is not sufficient; the addressee should also know how strongly she is being obliged to perform the action. It must be mutually known between speaker and addressee what the speaker's right is to request this action. Herrmann (1983) calls this the *legitimacy* of a request. A soldier who wants his trousers ironed by his colonel will probably have a hard time transmitting the legitimacy of the request to the colonel. It must also be mutually known that the addressee is able and willing to perform the action. The addressee's willingness will depend, in part, on the legitimacy of the request. Hence, the information to be conveyed by means of a request involves at least the following points: The speaker wants, with a certain degree of firmness or commitment, a particular action on the part of the addressee; the speaker has a certain legitimacy to oblige the addressee to this course of action, and is aware of the latter's ability and willingness to perform the action.

It is, however, a rarity that all information to be conveyed in a request is also explicitly expressed. Given the current discourse model and the

common ground of the interlocutors, a speaker can economize substantially on what is to be said. The addressee will infer the implicated information. What the speaker will select for expression depends on the firmness of his wanting the action, his feeling of being legitimized to request the action, and his estimates of the partner's willingness and ability to perform the action.

Herrmann (1983) and his co-workers manipulated some of these factors in experimental studies of requesting. One experiment consisted of a two-person game played by an accomplice of the experimenter and the subject, whose speech was analyzed. The players were in the role of detectives who had to perform certain tasks. The subjects were 144 young German adults. The game was organized in such a way that at some moment the subject couldn't do anything else than request the partner to hand over his pistol. It was, however, mutually known that the partner also needed the pistol himself; the willingness of the partner was not self-evident. What Herrmann varied in this experiment was the degree to which the speaker could view himself as being legitimized to request the pistol. The pistol was either his own property or his partner's. This factor strongly affected the speakers' choice of request form. When a speaker owned the pistol, he perferably used the German equivalent of a request form such as *give me the pistol*, *you must give me the pistol*, or *I need the pistol*. These forms do not express any concern about the addressee's willingness or ability to hand over the weapon. The first two forms, moreover, express the addressee's obligation to act, whereas the third one formulates the firmness of the speaker's want. If, however, the partner owned the pistol, the speaker was more likely to say *could you give me the pistol*, *would you give me the pistol*, or *I would like to use the pistol*. The first two of these express concern for the listener's condition, leaving a loophole for the addressee to deny fulfillment. The third form expresses the want but not its firmness, nor does it specify the action or any obligation to perform it.

In other experiments, Herrmann (1983; see also Winterhoff-Spurk, Herrmann, and Weindrich 1986) varied the perceived willingness of the partner to comply with the request, and found that speakers tended to express their concern for the partner's conditions of willingness or ability when these were in doubt. However, this was seldom done when the speaker was in a strong position of legitimacy to make the request. Politeness increased with diminishing rights. The rule is, apparently, that the information to be expressed should be two-way instrumental. It should, on the one hand, be sufficient for the addressee to derive the intended obligation; i.e., she should be led to believe not that the utterance is *just* an

assertion, a question, or an expressive but that it indeed conveys a request. The utterance should, on the other hand, minimize the risk that the addressee will decline to comply because bad feelings have been provoked. The latter risk is small when the speaker's right to oblige is very evident.

Also, Francik and Clark (1985) found that speakers construct requests in such a way as to overcome potential obstacles on the part of the addressee. When a speaker is put in a position to ask the time of somebody who is clearly not wearing a watch, the request tends to be of the form *Do you have a watch?* If, however, the addressee does wear a watch, the typical request is *Do you know what time it is?*.

So far, we have considered two cases of a speaker's selecting information for expression: in referring to objects and in making requests. The experimental results showed that, as a rule, speakers do not utter all the information that is to be conveyed. Rather, they select for expression information that is instrumental in achieving the communicative goal. In referring, this is the information that will be effective in focusing the addressee's attention on the new referent, be it endophoric or exophoric information. In requesting, it is information that will minimize the risk of a noncomplying response on the part of the addressee; this, in turn, depends on the distribution of rights, the perceived willingness of the addressee, and so forth.

### 4.3.5 Selecting Main-Structure and Side-Structure Information

The problems of information selection are substantially more complicated when we consider complex verbal activities, such as giving route directions, describing scenes, narrating, and planning joint actions. In these and other cases, a main goal is successively unfolded in subgoals and sub-subgoals. We will return to some of these issues in the next section; here we will note one rather general feature of a speaker's selection of information—a feature that is common to almost all these complex types of discourse: Speakers categorize the information they select for expression in what Klein and Stutterheim (1987) call *main-structure* and *side-structure* information. For instance, one speaker, when asked to relate her plans for the future, said (among other things) the following:

(5) So, I will go to the university and study something, probably French. And then, I will become a teacher, *although the chances are bad right now.* And then of course, I will marry and have children. *I am very traditional here, I love babies.*

The main structure of this text consists of the subsequent future steps. They are straightforward elaborations of the main goal the speaker set for herself

in response to the interviewer's question. The side structure (italicized in the text) consists of additional comments, associations, embellishments, and so forth. Hopper (1979) and Reinhart (1984) have called the latter *background* information and the former *foreground* information. It is, in general, much easier to see how the selected main-structure or foreground information relates to the speaker's communicative intention than to explain why the background or side-structure information is selected for expression. The speaker can generate all sorts of "side intentions," such as to give reasons for actions, plans, or decisions; or he can develop non-communicative intentions (in the sense discussed in section 4.1), such as to appear knowledgeable, pleasant as a conversant, etc. Whatever the grounds are for the selection of side-structure information, the speaker gives it special treatment in the generation of the message. For instance, it is not given the same temporal deictic perspective as the main-structure information. In example 5, the future tense of the main structure does not appear in the side structure. Also, the side messages express states rather than events, and there are various other remarkable differences (see Klein and Stutterheim 1987).

In one experimental study in which the linguistic differentiation of main-structure and side-structure information in speech production was examined, Brown and Dell (1987) asked subjects to retell stories. Each story involved an action performed by means of some instrument (e.g., a stabbing). One experimental variable was the kind of instrument used. It could be the typical instrument for the action (a knife) or an atypical instrument (an icepick). When speakers retold such a story, the typical instrument was less often explicitly mentioned than the atypical instrument. Since the default instrument for stabbing is a knife, the speaker can refrain from selecting that information for expression; it is inferable. When a speaker nevertheless decides to mention the knife explicitly, this must be due to a "side intention"—to embellish the story, to be very explicit, or what have you. The atypical instrument, however, is main information to be expressed. Brown and Dell found a characteristic difference in the grammatical encoding of atypical and typical instruments. Main information (i.e., the atypical instruments) tended to be encoded in the same clause as the action (*The robber stabbed the man with an icepick*), whereas side information (i.e., the typical instruments) tended to end up in an earlier clause (*The robber grabbed a knife and stabbed the man*). In other words, the speaker's message expressed a more intimate relation between action and instrument when the instrument belonged to the story's main structure than when it belonged to its side structure.

## 4.4 Macroplanning 2: Ordering Information for Expression

Whenever a speaker wants to express anything more than the simplest assertions, requests, declarations, etc., he has to solve what will be called the *linearization problem*: deciding what to say first, what to say next, and so on. The linearization problem has been the subject of rhetorical treatises since Aristotle; an educated speaker was supposed to give special attention to the ways in which he would order or arrange information for expression. That such arrangement (even of two simple propositions) can have dramatic effects on the addressee's interpretation is apparent from the following examples:

(6) She married and became pregnant.

(7) She became pregnant and married.

Each of these sentences contains the same two propositions; however, their order differs, and the implicatures differ correspondingly.

It is important to distinguish the linearization of propositions from the kind of ordering that is due to topicalizing or taking perspective in a message. (The latter is called "linearization" in Chafe 1970.) The two sentences in example 8 differ in this respect; the second one is "topicalized."

(8) I will send you the money next week.
The money I will send you next week.

Such topicalizing effects will be discussed in subsection 4.5.3. Here our main focus is the linearization of entire propositions or predications.

There are two major sets of determinants for a speaker's ordering of information for expression: *content-related* and *process-related* determinants.

### 4.4.1 Content-Related Determinants

The content-related determinants of linearization derive from the following principle:

*Principle of natural order* Arrange information for expression according to the natural ordering of its content.

What counts as natural ordering is different for different domains of discourse, and there is no general definition. Still, for certain important cases the notion is obvious. For event structures, the natural order is the chronological order of events. Unless the speaker explicitly indicates otherwise, the interlocutor will assume that the order of mention corresponds to chronological order. This is what happens in examples 6 and 7 above; it is the default case. A cooperative speaker will mark deviations from chro-

nology, as in the following:

(9) She became pregnant after she married.

The word *after* signals to the addressee that chronology is not preserved. It is known that children have a hard time acquiring these deviant structures (E. Clark 1970); preserving chronological order is one of the earliest rhetorical skills in children.

There are also other domains that have natural ordering. Linear spatial structures are a good example. Above, we considered a speaker's retrieval of the shortest or easiest route from a source place to a goal place in town. It is not enough for a speaker to make this spatial information conceptually available; the information should also be ordered for expression. The natural order is the connective sequence of loci from source to goal. It is not necessarily the case that this is also the order in which a speaker retrieves the shortest route from memory. He may well happen to become aware of the final part of the route before he has worked out the initial part in detail. In other words, the natural order has to be imposed for the listener's sake. That speakers indeed follow source-to-goal spatial connectivity when giving route directions is evident from several empirical studies (Klein 1979, 1982; Munro 1977; Wunderlich and Reinelt 1982).

Why is natural order so natural? The principle of natural order will be effective only in domains of discourse where there is tacit agreement between the interlocutors as to what will constitute a natural order. This tacit agreement may be due to universal principles of memory organization, or to more culture-specific "scripts" (Schank 1975). An example of the former is, probably, the chronology principle for temporal domains. If people normally organize and remember related events as temporally ordered structures, it should be relatively easy or natural for a speaker to retrieve the information, and for a listener to decode it, in that order. A *script* gives a further specification of what the order should be in a particular culture. The order of courses in a meal, for instance, is not universal; within a culture, however, there will be a script, shared by the language community, which can provide a set of default values for arranging messages about meals.

### 4.4.2 Process-Related Determinants

The process-related determinants of linearization are most apparent when there is no natural order, and especially when the speaker has to express a multidimensional informational structure. A classical example is the task of describing the layout of one's apartment (Linde and Labov 1975). Two-

or three-dimensional spatial information has to be mapped onto some linear order for expression, and there is no single natural way to proceed. The major limiting factor is the speaker's bookkeeping ability. The speaker must keep track of what has been said and what is still to be expressed, and this requires special memory operations. Also, the speaker may show some concern for the memory limitations of the addressee, who will have to reconstruct an image of the multi-dimensional array from the successive bits of information.

Results from experimental studies of the process-related principles of linearization are reported in Levelt 1981, 1982a,b, and 1983. Subjects in the experiments were asked to describe spatial-grid-like networks, such as those in figure 4.5, which were put on the table in front of them. These networks consisted of differently colored dots, connected by horizontal and vertical lines. (In the figure, the colors are replaced by their names.) The subjects were asked to start their descriptions at the node indicated by an arrow, and to describe the network in such a way as to enable the next subject to correctly draw it on the basis of the tape-recorded description. Here is one subject's description (translated from Dutch) of pattern a:

(10) Begin in the middle, a gray node. From there upwards a red node. Then to the left, a pink node from the red. Then from pink again to the left a blue node. Then back again to red. Then from red to the right a yellow node. And from yellow again to the right a green node.

The precise models that successfully predicted the subjects' linearizations (Levelt 1982a) are less important here than the principles on which they are based. It is likely that these principles will, *mutatis mutandis*, hold for other spatial and nonspatial domains of description, because they reflect quite general properties of perception and memory. Let us review them with reference to the various patterns in figure 4.5.

*Principle of connectivity* Wherever possible, choose as the next node to be described one that has a direct connection to the current node.

This principle predicts that a speaker will go over a pattern as much as possible "without lifting the pencil," the mental pencil's point being the speaker's focus of attention. In example 10 above, the speaker goes in a connected fashion from GRAY to RED to PINK to BLUE, and again from RED to YELLOW to GREEN. Speakers rarely violated the connectivity principle for string-like parts of patterns. Nobody ever went from RED to GREEN to YELLOW in pattern a. For pattern b, every subject went in connected fashion from GREEN to PINK, mentioning all intermediary nodes as if giving route directions.

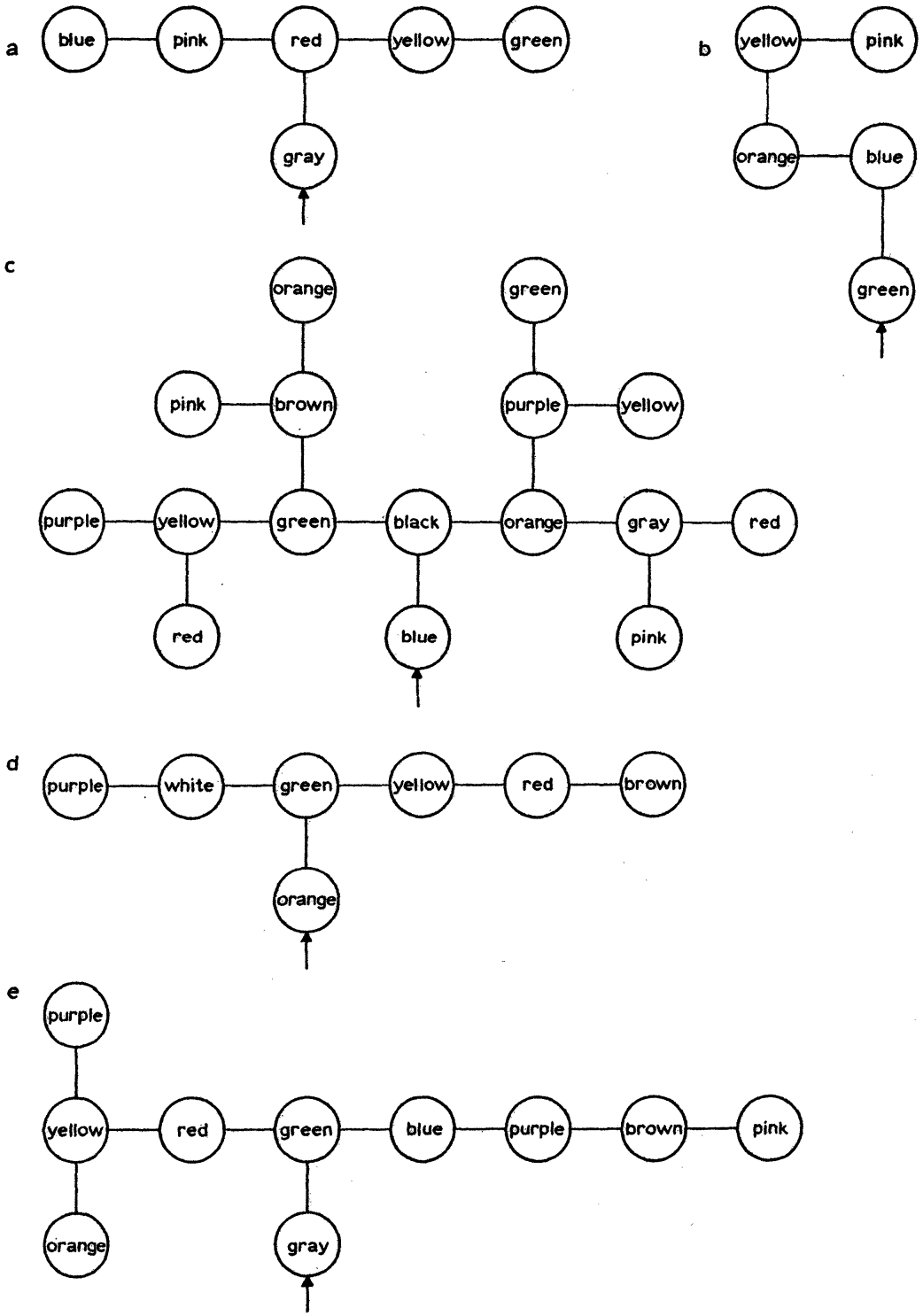



**Figure 4.5**
Patterns to be described in a linearization experiment (After Levelt 1982a).

Ehrich and Koster (1983) found a high degree of connectivity in the description of furniture arrangements in a doll house. This was especially marked when the arrangements were chaotic or nonfunctional; the speaker would make small jumps from one object to the nearest one. There was far less connectivity of this sort when the furniture arrangement was natural (e.g., chairs around a table). In that case, natural order took over; a speaker would introduce the central object (the table) and mention all objects with some functional relation to the table (the chairs, the lamp, the vase).

The connectivity principle is certainly not limited to spatial domains of discourse. Whenever possible, a speaker who is asked to list his relatives selects as the next person to mention one who has direct kinship to the last person mentioned (Levelt, unpublished data). One goes from ME to PARENT to UNCLE or AUNT to COUSIN, never straight from PARENT to COUSIN. Connectivity is a general ordering principle in perception and memory.

It is, of course, not always possible for a speaker to introduce new nodes (new items of information) in a connected way without repeating old ones. This problem appears when the speaker reaches the end of a string in patterns such as the ones in figure 4.5. In example 10, the speaker leaps from the end point (BLUE) back to the choice node (RED) in order to complete the description of the right part of the pattern. The return leap violates connectivity, but two-thirds of the subjects do this. The only way to preserve connectivity would be to return step by step to the choice node, retracing the old items. This is what one-third of the subjects do. Here is an example description for the same pattern (figure 4.5a):

(11) I start at node gray. Go straight on to red. Go left to pink. Go straight on to blue. Turn around, go back to pink. Go back, uh straight on to red. Straight on to yellow. Straight on to green.

The speaker preserves connectivity by adding that one should "turn around" and by rementioning PINK. The speaker is, clearly, involved in a mental tour through the pattern. Repeating old items preserves connectivity but violates a quantity maxim. Most subjects opt to respect the latter.

How do "leaping" subjects keep track of the returns to be made? This is a simple matter for pattern a, where there is only one choice node to leap back to. The speaker mentally flags "back to RED upon reaching the end of a string." But if the speaker passed two or more choice nodes before reaching the end of a string, in which order should he return to these nodes? This problem arises when pattern c is described. When the speaker starts his description by mentioning in connected fashion BLUE, BLACK,

ORANGE, PURPLE, and GREEN, there are three choice nodes in the queue for the speaker to return to: BLACK, ORANGE, and PURPLE. The next principle expresses how speakers handle this.

*Stack principle* Return to the last choice node in the waiting line.

The just-mentioned speaker who reached GREEN will first return to PURPLE, and will mention the YELLOW node to the right of PURPLE. He will then go back to ORANGE and mention GRAY, which is itself a choice node. After mentioning RED (or PINK), the speaker will return to GRAY, the last choice node in the line, in order to mention PINK (or RED); only then will he return to BLACK. The speaker's bookkeeping for return addressees is like putting them on a stack and always returning to the top item on the stack after reaching the end of a connected string. The experimental data show almost no exceptions to this principle.

This stack principle is well known in the psychology of problem solving (see, for instance, Newell and Simon 1972); it is a dominant way of keeping track of hierarchically organized structures. It is not unlikely that speakers will also follow the principle in linearizing other spatial or nonspatial domains that are multiply branching. Linde and Labov's (1975) findings for apartment descriptions are in full agreement with the principle.

The third and final principle regulates what speakers will do upon reaching a multiply branching node. In which order do they treat the outgoing branches? Consider pattern d in figure 4.5. Which branch will the speaker describe first upon reaching the GREEN choice node: the left branch, or the right one? There is a probabilistic rule governing such choices:

*Minimal-load principle* Order alternative continuations in such a way that the resulting memory load for return addresses is minimal.

For pattern d, this principle predicts that speakers will tend to describe the left branch first. The speaker will have to keep the GREEN choice node flagged in memory for return, whether he goes left or right first. But if he goes left first, the duration of the load will be shorter; there are only two further nodes before the return leap can be made. When the speaker goes right first, as many as three nodes have to be described before memory can be relieved of its flag. And indeed, speakers mostly describe shorter branches before longer branches.

Another example is pattern e. Which branch will be taken first from the GREEN choice node? Both branches contain the same number of nodes and arcs. Still the principle predicts a preference for going right first. There are two choice nodes in the pattern: GREEN and YELLOW. When the speaker goes right first, he will keep GREEN on the stack until he reaches

PINK. But after the return leap to GREEN, the stack is empty until YELLOW is reached. There is, in other words, never more than one return address to be remembered at a time. However, when the speaker goes left first, he will have to flag GREEN and (after two more moves) YELLOW as well; he will then have two return addresses on his stack. Hence, the memory load will be minimized by going right first. And that is indeed what speakers prefer to do in the experiments.

When there are more alternatives, the principle says "Do the simplest thing first." And one can expect this to apply to other domains of discourse as well. When one has to give instructions for shopping involving two different shops, a big food store where a large range of items has to be purchased and a dry cleaner where a skirt has to be picked up, one will probably start the instructions by mentioning the dry cleaner. But only further research can tell us how general this principle and the previous two are as process-related determinants of linearization.

This section and the previous one have discussed aspects of macroplanning, the processes by which a speaker selects and orders information for expression. The result of marcoplanning is a speech-act intention, or a series of speech-act intentions. The speaker selects and orders information whose expression with declarative, interrogative, or imperative mood will be instrumental in realizing the goals that proceed from the original communicative intention. In other words, macroplanning produces the substance of the messages, such as that the message should declare a particular proposition or interrogate a certain state of affairs. But more has to be done in order for a message to become expressible. Its contents must be presented in such a way that the addressee can attend to it, find the referents, register what is new, and so forth. Also, the contents must be put in a propositional format that the Formulator can understand. This is microplanning, to which we now turn.

## 4.5 Microplanning

Four major aspects of microplanning will be discussed in this section. The first one is the assignment of what I will call *accessibility status* to the referents in the message. Each referent in the message will be provided with an index which states where it can be found—for instance, in the current focus or elsewhere in the discourse model. This index will be taken into account in grammatical encoding, giving rise to cues in the utterance that guide the addressee's attention to where the referent is to be identified. The second aspect of microplanning to be treated is topicalization, the assignment

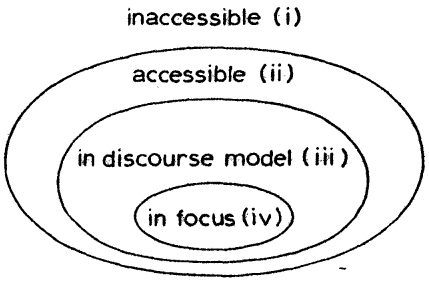
of the topic role to one of the referents. The third is the "propositionalization" of the information to be expressed. It will be argued that translating information into propositional format necessarily involves the assignment of perspective. The final aspect of microplanning to be discussed is the acknowledgment of the Formulator's language-specific requirements.

### 4.5.1 Assigning Accessibility Status to Referents

A speaker will introduce and reintroduce referents (persons, objects, events, etc.) into the discourse in such a way that the addressee can create or locate them in her own discourse model. What matters here is the estimated *accessibility* of the referent for the addressee ("estimated" because it is the speaker's judgment that matters, not the real accessibility as experienced by the addressee). Recall from section 4.2 the conversation between Marcia and Seth. Marcia had, at some point, said something about the smoke of Florence. Owing to a lapse of attention, Seth had not registered that remark. When Marcia, not aware of this, wants to say more about this smoke, she will judge that it is accessible as a referent for Seth. She may then refer to it as *the smoke*. But Seth, who cannot locate the referent in his discourse model, will ask *what smoke?* or something similar.

The accessibility status of a referent can be conceived of as a complex value or index attached by the speaker to each referent in a message. This value is an important determinant of the linguistic shape the Formulator will compute for the referent. Marcia's discourse entity SMOKE may surface in her utterance as *the smoke*, *smoke*, *that*, or *it*, depending on what she thinks is the entity's accessibility to Seth.

The speaker can derive the accessibility status of a referent from his content and focus accounts (see subsections 4.2.3 and 4.2.4). This involves three dichotomies, which are of central relevance for further grammatical encoding: whether the referent is estimated to be accessible at all to the addressee, whether it is assumed to be in the addressee's discourse model, and whether it is guessed to be in the addressee's current focus. These three dichotomies are not orthogonal. They are, rather, embedded in this way:



This embedding of dichotomies naturally corresponds to four cases, numbered in the diagram. For different languages, the consequence for grammatical encoding of a referent's accessibility status on these dichotomies may be quite different. But it seems that most languages acknowledge these dichotomies in some way or another. In discussing the four cases, I will give examples for English only. The examples cannot be simply generalized to other languages.

(i) The discourse record tells the speaker that the listener cannot find or infer the entity. It is neither in the common ground nor in the discourse model; it is also not inferable (see (ii) below) from entities in these shared knowledge structures. Prince (1981) calls these referents "brand new. ' The ferry Marcia wants to introduce in the discussion (see figure 4.2) is not represented for Marcia as something Seth knows about. The English speaker will normally encode a referent with the feature "inaccessible" as indefinite. For instance:

One can reach the island by an old ferryboat.

Marcia would not be very cooperative if she used a definite expression, as in

One can reach the island by the old ferryboat.

Marking a referent as inaccessible is inviting the listener to create a new address for the entity referred to.

(ii) Though the referent is neither in the discourse model nor in the common ground, the discourse record makes it seem likely to the speaker that the addressee can *infer* the referent. Marcia believes that she has successfully introduced the information that Ponza is an offshore island, one of the Pontines. She also believes that Seth is focusing on this information. Marcia may now rely on Seth's general knowledge about Italian offshore islands, and in particular that such islands tend to have old ferries. Seth, she thinks, can infer that Ponza has such a ferry. She may then say

One day I took the old ferryboat.

It is essential that Ponza, from which the existence of the ferry can be inferred, is supposed to be in the addressee's focus. When Marcia believes that Seth is focusing on Florence, she cannot refer to Ponza's ferryboat by *the old ferryboat*. By using a definite expression, a speaker tells an addressee that a referent can be uniquely identified—either in their common ground (Seth could talk about *the pope*), or in their shared discourse knowledge, or by inference from what is currently in focus. In the last case, the listener will create a new address by backward suppletion.

The following two cases concern reference to entities that are already in the discourse model. When the entity is uniquely identifiable, the speaker will make definite reference to tell the addressee "you can identify the referent". But the speaker can do more to guide the addressee's attention. He can make nonprominent or reduced reference.

(iii) The entity is in the discourse model, but it is not in the addressee's current focus. As in the previous case, such a referent is uniquely identifiable; it receives the index "accessible". The English speaker will express this by making definite reference. But in addition to this, the referent is indexed for being available in the discourse model. It has, in that sense, no news value. When an entity is newly introduced into the discourse it does have news value. This notion, called *conceptual prominence*, will be discussed in the next subsection.

An item already in the discourse model, and not prominent for other reasons, will receive prosodic deaccentuation in the (English) utterance. Pitch accent will be withheld. When Marcia talks about Florence after relating her experiences at Ponza, she may, as an afterthought, want to add something about Ponza's church:

The island [or Ponza] has a beautiful chUrch.

The word *island*, or *Ponza*, whichever is used, can be uttered in a nonprominent way. Accentuation goes to the referent that is *outside* the discourse model: Ponza's church.

The speaker's tendency to deaccent the expression for a referent that is already in the discourse model has been revealed in various empirical studies. It is, for instance, in excellent agreement with the findings of Brown and Yule (1983), who instructed speakers to describe visual diagrams containing colored geometrical figures, words, and connecting colored lines. Each speaker had as addressee a person who could not see the diagram, but who was to draw it from the speaker's description. Here is a typical description:

halfway down the page draw a red horizontal line of about two inches on eh the right hand side just above the line in black write "ON".

When the tapes were analyzed for prosodic prominence or nonprominence of referential expressions, the results shown in table 4.1 emerged. These clear findings are also in good agreement with those of MacWhinney and Bates (1978), Marslen-Wilson, Levy, and Tyler (1982), Terken (1984), and Fowler and Housum (1987). Fowler and Housum compared speakers' utterances of words produced for the first and for the second time in a monologue (both times referring to the same entity). It turned out that

**Table 4.1**
Prominence and nonprominence in diagram descriptions (data from Brown and Yule 1983).

| | Referent is | | | |
|---|---|---|---|---|
| | not in discourse model | | in discourse model | |
| | inaccessible | inferable | not in focus | in focus |
| Prominent | 87% | 79% | 4% | 0% |
| Nonprominent | 13% | 21% | 96% | 100% |

speakers distinguished old referents by attenuating their names. These words were shorter, lower in pitch, and less loud at second use.

Let us now return to Marcia, who said *The island has a beautiful church.* Notice that, under the same circumstances, she should not say *It has a beautiful church*; the pronoun *it* would have been interpreted by Seth as referring to Florence, which is in Seth's current focus. This brings us to the fourth case.

(iv) The referent is in the addressee's current focus. Such a referent is given the status feature "in focus". Because of the embedding relation depicted above, each "in focus" referent is also in the discourse model and accessible. This means that the English speaker will, normally, deaccent the referring expression and make it definite. But the specific "in focus" feature has an additional consequence: The referring expression will be reduced. When Marcia believes Seth is focusing on the island of Ponza, she can say *It also has a beautiful church* or *I have seen a beautiful church there.* The proforms *it* and *there* tell the addressee "Don't search; you have the referent in focus." This use of reduction has been especially well documented by Chafe (1976), Prince (1981), Marslen-Wilson, Levy, and Tyler (1982), and Brown and Yule (1983). But it should be added that there are degrees of reduction (or "attenuation," as Chafe called it). Marcia would probably not say *The island of Ponza has a beautiful church,* which contains a very full referential phrase (*the island of Ponza*). Less full are *Ponza island, the island,* and *Ponza*; very reduced are *there* and *it.* Speakers tend to reduce the size of referential expressions when an entity is repeatedly referred to in discourse. Empirical evidence to this effect has been reported by Krauss and Weinheimer (1964), Osgood (1971), Brown and Yule (1983), Clark and Wilkes-Gibbs (1986), Redeker (1986), Isaacs and Clark (1987), and Sridhar (1988). Adjectives, relative clauses, and other modifiers tend to be dropped. Interlocutors follow Grice's maxim of quantity by economizing on the size of referential expressions. This economizing is also done for discourse entities

that are not in focus. However, when a referent is believed to be in the addressee's focus, and especially in her focal center, reference is, as a rule, made by the use of "pro-forms," such as *it*, *he*, *that*, and *this*.

These four cases illustrate a referent's status on the three accessibility features "accessible", "in the discourse model", and "in focus". The speaker's assignment of this three-valued index to each referent can be seen to be governed by procedural knowledge. The underlying procedures can be put this way:

$$\text{IF the referent} \left\{ \begin{array}{l} \text{can be uniquely identified} \\ \text{is in the discourse model} \\ \text{is in the addressee's current focus} \end{array} \right\},$$

$$\text{THEN assign it the value} + \left\{ \begin{array}{l} \text{accessible} \\ \text{in discourse model} \\ \text{in focus} \end{array} \right\};$$

otherwise assign it the opposite (negative) value.

The English encoding of these three features employs, in particular, definiteness, deaccentuation, and reduction (pronominalization), respectively. Other languages express these features differently, and even in English there are known exceptions to these rules of thumb. One of the exceptions has to do with the assignment of prosodic prominence: An item already in the discourse model is not necessarily subject to deaccentuation; it may have conceptual prominence for independent reasons. We will turn to this now.

### 4.5.2 Conceptual Prominence

When a speaker introduces a new referent into the discourse, it has what we called "news value" for the addressee; it is conceptually *prominent*. But a referent that is already in the discourse may also become prominent. Consider the following example.

Assume that Seth and Marcia shift to a new topic: a quarrel between three boys and a girl they know, Sam, Saul, Simon, and Tessie. After introducing the protagonists, Seth could say

First, Tessie pestered SAm, then she insulted SAUl, and finally she hit SImon.

All three victims (i.e., patients of some action) are given pitch accent here in spite of the fact that they are all represented in the discourse model. Nooteboom and Terken (1982) pointed out, and showed experimentally, that pitch accent gets assigned to successive but different referents appear-

ing in the same role. In our example this is the case for Sam, Saul, and Simon, who are all figuring as patients in Tessie's actions. (Pechmann's [1984] experimental evidence to this effect was discussed above.) One could say that in these cases the referents *contrast* in the same thematic role (in Pechmann's terms: there is endophoric contrast). The news value is in the fact that the same role, the one in the listener's focus of attention, is now fulfilled by a different referent. In order to express this, the speaker will mark it in the message by assigning prominence to the item. This, in turn, may lead to accentuation of the referential expression.

When, however, the *same* referent reappears in the *same* role, it is made nonprominent. In the example that happens to Tessie, who continues to be actor. Whatever the prominence of the first mention, the second and third (both *she*) are deaccented.

Since repeated mention is usually made by anaphoric means—that is, by terms that stand for the original referential expression (e.g., *she* for *Tessie*) —one would expect anaphors, and in particular anaphoric pronouns, to become deaccented; they refer to entities already in the discourse model. But pronouns may also become accented when their referents contrast in the same conceptual role. Consider the following example:

Tessie went after SImon, and then hE chased hEr.

Here *Tessie* and *Simon* are reduced to pronouns (*her* and *he*), but these reduced forms receive pitch accent. It signals, for both referents, contrast in the same role (the actor was first Tessie, then Simon; the patient was first Simon, then Tessie). Here the reduction to pronouns tells the addressee that the referents are still in focus; the accentuation tells the addressee that the same kind of action is continued, but with the roles swapped. Note that the roles in the action must indeed be very similar if this contrastive prominence is to be used. The following could not be said:

Tessie went after SImon, and then hE kissed hEr.

Going after and chasing are similar in a way that going after and kissing are not. These cases show that speakers not only mark referents for being in the discourse model or not; they also mark them for contrastive roles between what is in the addressee's current focus and what is in the speaker's new focus.

The assignment of prominence to elements in a message is not restricted to referring items. Predications can also be marked as prominent by the speaker; this again depends on their estimated news value for the addressee, as is particularly apparent in question answering. Consider the following exchange.

Q: What happened to your uncle?
A: He dIEd last week.

The answer provides the new information, which the questioner presumably asked for. This new information is given prominence in the message. And conceptually prominent information is usually formulated with pitch accent, also when it is a predication.

In short, a conceptual entity in a message is assigned the feature "+ prominent" when it has news value. Though this notion is hard to formalize, three clear cases are recognizable: A speaker will refer with prominence when the entity is newly introduced in the discourse, i.e., when it has the accessibility feature "– in discourse model". He will also refer with prominence when the referent is contrasting in a focused role. And he will assign prominence to a new predication, especially when this is an answer to a question about some referent.

### 4.5.3 Topicalizing

When a speaker has selected certain information for expression, he will use various devices to guide the listener's attention. One such device, treated in the previous subsection, is to signal to the addressee the location of a particular referent or the need to add a new address to the discourse model. But there is more to be done (see section 3.4). It may often be necessary for the speaker to topicalize a referent.

A speaker will mark as topic the referent that the message is about. In this way the speaker may tell the addressee where to store the information being expressed. If the speaker intends to inform the other party about his sister Jane, he could say

Jane is married to Peter.

Here, *Jane* is topic, and the addressee is invited to store the new information (being married to Peter) under Jane's address in the discourse model. If there is also an address for Peter (i.e., when Peter is accessible inside or outside the discourse model), the listener may, in addition, store the corresponding information (Peter's being married to Jane) under Peter's address. However, the speaker probably intends only the former when he makes Jane the topic. In procedural terms:

IF the goal is that the listener store the information under address X,
THEN assign topic status to X.

Why this should become a goal is a different issue; it should, in some way, proceed from the speaker's illocutionary intention. One obvious reason for setting oneself such a goal is an interlocutor's explicit question, for instance

*Is your sister married?*. Another important reason for topicalizing a referent is that it is particularly *salient* in the state of affairs to be communicated. A human agent is more easily topicalized than a nonhuman theme (*Peter cut the tree* versus *The tree was cut by Peter*), big or moving objects are more easily topicalized than small or immobile ones, and so on. We prefer to entertain addresses for salient items; they are the pegs for storing our information about the world. Notice that as soon as these salient items have been introduced into the discourse model, the speaker will tend not to mark them for prominence any more. Instead, the predications made about them—that is, the new information items which the addressee is invited to hang upon these pegs—will be prominent. The relation between topicalization and saliency will be taken up again in chapter 7, where the grammatical encoding of topicalized and nuclear elements will be discussed.

In running discourse there is often a sequential relation between focusing and topicalizing. This is exemplified by the following:

I have a sIster. She is mArried.

The first sentence brings a new referent into the addressee's focus of attention: the speaker's sister. Since it is as yet inaccessible to the hearer, the speaker uses an indefinite full form of reference with prosodic prominence. In this way the listener is invited to set up a new address in her discourse model. The next sentence has this referent as topic. It is now in the addressee's current focus, and it is therefore referred to in short pronominal form (*she*) and without prosodic prominence. Now the predication receives prominence. It often takes two subsequent steps to introduce a referent and to say something about it: focusing and topicalizing.

When a referent is topicalized in a speaker's message, it will be given a kind of priority treatment in grammatical encoding. It will, for instance, tend to be expressed as the grammatical subject. Entities that are not topicalized but are still quite salient (i.e., the other nuclear entities of the message) will also be given special treatment; they will tend to be encoded in a major grammatical function. We will return to these issues in chapter 7.

### 4.5.4 Assigning Propositional Format and Perspective

A speaker often decides to communicate information that is not yet in propositional format. Take again the case of giving route directions. A speaker, when asked to tell the way to the Concertgebouw, will do the macroplanning that will give him the successive moves to be made by the addressee. But each of these moves will, initially, not be in propositional

form. Rather, these moves will be spatio-temporal images. An important aspect of microplanning is to translate these images into propositional form. The main point of this subsection is that this translation necessarily implies the assignment of perspective.

The notion of perspective-taking is easily exemplified by descriptions 10 and 11 above, which were given by two subjects in the experiment on linearization mentioned (Levelt 1982b). The descriptions refer to the pattern in figure 4.5a. Let us consider a single move in these descriptions, namely the one concerning the connection PINK-BLUE. The first subject, who gave description 10, put it this way: *Then from pink again to the left, a blue node.* The second subject, giving description 11, said: *Go straight on to blue.* Apart from the difference in ellipsis, the descriptions show a contrast in the directional terms used: *to the left* in example 10, *straight on* in example 11. Although the perceptual structures were identical for the two subjects (a pink node and a blue node connected by a black horizontal arc), they were categorized in different ways; the spatial relation of BLUE to PINK is TO THE LEFT OF in example 10 and STRAIGHT ON in example 11. The speakers agreed on their choice of PINK as the reference location for BLUE; nobody in the entire experiment said *Pink is to the right of blue.*

Both the commonality and the difference observed reveal the workings of perspective-taking. That all subjects relate BLUE to PINK and not the other way round is a consequence of the linearization strategy they have taken, namely to describe the pattern in a connected fashion as if making a tour. After the speaker has reached and mentioned the pink node, the continuation of the tour requires him to relate the next node to the last one mentioned—i.e., BLUE to PINK, since PINK is the current focus. There can be no doubt that all the speakers in the study would have related PINK to BLUE if they had been asked to start the entire pattern description from the BLUE node. When a speaker is not bound by instruction and neither PINK nor BLUE is currently in focus, he can freely take either perspective, and his categorizations of the binary spatial relations will be dependent on his choice of reference point. Note, however, that *some* choice is to be made; either BLUE or PINK is to be the relatum for putting the spatial relation between BLUE and PINK in propositional form.

The difference between the two subjects is a subtle one that reveals another degree of freedom in perspective-taking.

The first subject describes the pattern in terms of deictic perspective, taking himself as the basis for the coordinate system (see subsection 2.2.2). The mental tour made by the subject is a gaze tour; all directional terms tell the addressee where the gaze moves in terms of a speaker-centered base.

The speaker, who examines the pattern on the table in front of him, has to move his gaze to the left in order to go from the pink node to the blue node. Strong evidence that speakers who say *left* here apply categories of gaze movement to the pattern is that almost all these speakers also say *up* or *upwards* in describing the first move from the gray to the red node, *never straight* or *ahead.* The gaze's move is indeed upward for this first step in the tour, in spite of the fact that the pattern is flat on the table. In Levelt 1982b I proposed a gaze-tour theory for the frequent use of vertical dimension terms (*up, down,* etc.) for spatial relations that are essentially in a horizontal plane. The principle was later also put forward by Shepard and Hurwitz (1984). In short, speakers who take a deictic perspective choose *categories of gaze movement* (UP, DOWN, LEFT, RIGHT) as functions for locative predications.

The other speaker, who said *straight on,* doesn't make a gaze tour from a fixed deictic perspective; rather, she makes a kind of body tour, as if driving through the pattern by car (remember her mention of "turning around"). For every move through the pattern, the direction of the line moved along previously is taken as the new intrinsic basis for the coordinate system. It is reoriented as moves are made. The *straight on* derives from just having driven from the red node to the pink node, and continuing the same direction of movement. And indeed, in her description this subject uses *straight on* for all cases where the same direction is continued. Notice also that, as it should be in a flat array of streets, there are no *ups* or *downs* for the body-tour subjects.

This example shows that the same visual pattern can be propositionally expressed in quite different ways, depending on the deictic or intrinsic perspective taken. It determines which entities are reference locations for which other entities, and it determines the direction of these relations. Most subjects in the experiment preferred to take deictic perspective. But one-third of the speakers used the body-tour strategy, which asks for a pattern-based intrinsic perspective (the factual orientation of the speaker being irrelevant). This is, apparently, a matter of cognitive style. Is this a trivial difference? One should stay on the alert; it turned out that subjects with left-handedness among their parents or siblings were less inclined to take deictic perspective than other participants in the experiment. One wonders whether hemispheric lateralization plays a role in perspective-taking.

The example made it especially clear that there is no single *necessary* way of assigning propositional format to the visual information (and this was confirmed in a similar experiment; see Garrod and Anderson 1987).

One would expect such necessity if the visual system's output were itself in propositional form. In that case the perspective inherent in the proposition would simply be forced upon the mind. There is, however, substantial freedom in putting the perceived structure, which is spatially represented, into one or another propositional form. They are all *equivalent* descriptions of the same perceptual pattern. But they differ in perspective. This does not mean that perceptual factors play no role in assigning perspective. They do. Speakers prefer to express figure/ground relations such that the ground is taken as reference. They tend to say *The cat is in front of the wall* rather than *The wall is behind the cat.* What is smaller is preferably located with respect to what is larger. A chair-and-table scene is preferably described as *A table and a chair next to it* rather than *A chair and a table next to it.* When there is a moving object and a stable object, the moving object (A) tends to be located with respect to the stable one (B): *A approaches B,* or *A passes by B.* A contained object tends to be located with respect to the container: *A triangle in a circle* rather than *A circle around a triangle.* Rosch (1975) has argued that there are natural reference points not only in perceptual categories but also on such an abstract dimension as that of natural numbers.

We will finish this section by considering one empirical study of perspective-taking in the description of spatial scenes. Ehrich (1985) asked subjects to describe the arrangement of furniture in a doll house and analyzed the factors that determined the choice of reference objects and relations. These choices *must* be made when a spatial scene is expressed in propositional format. The study shows the systematicity of some of these choices. Ehrich varied a number of factors: the size and relative position of the objects, their locations and orientations with respect to the speaker, and the presence or absence of a background wall. But the subjects were left completely free in their choice of words.

Ehrich presented her subjects with furniture arrangements such as the ones in figure 4.6. Each subject inspected a single room through the door opening. In one condition the room had no walls. This, however, had very little effect on the speakers' descriptions; it can be ignored for the present purpose. The subjects examined the room with its six objects for some time, playing an "I spy" game with the experimenter. After this the room was removed, and the subject was unexpectedly asked to describe the furniture arrangement in such a way that a listener could reconstruct it. There were four major results:

(i) Speakers, in large majority, took a deictic perspective. In the case of the arrangement on the left in figure 4.6, they preferred to say *In front of the*

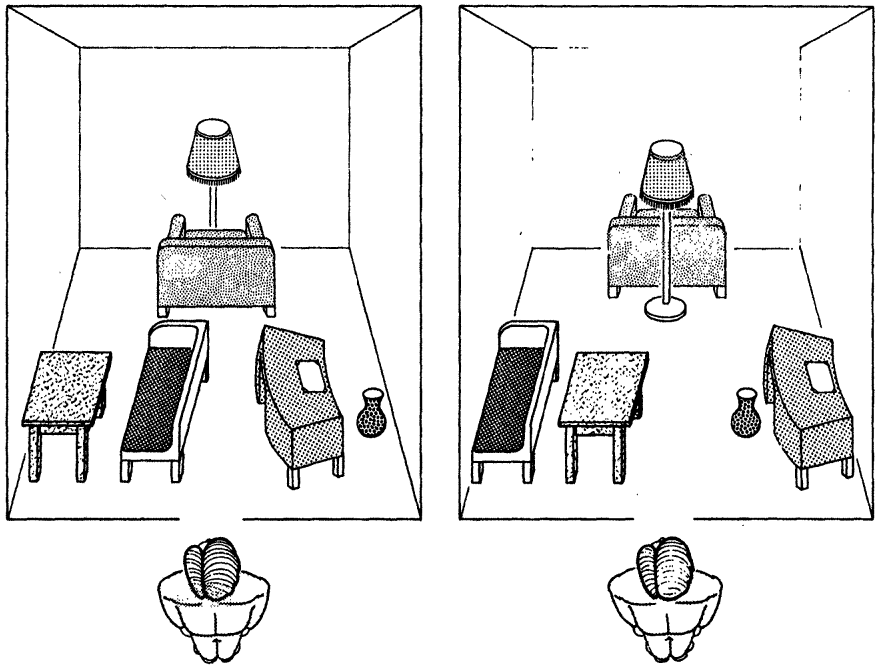


**Figure 4.6**
Examples of furniture arrangements to be described. (After Ehrich 1985.)

*lamp is a chair*, not *In front of the chair is a lamp*. In the case of the other arrangement, they even preferred to say *There is a lamp in front of the chair*, though the lamp was intrinsically behind the chair.

(ii) Speakers preferred to use the more peripheral object as reference for the location of an object that is closer to ego. In describing the arrangement shown at left, most speakers located the chair with respect to the lamp; in describing the other arrangement they located the lamp with respect to the chair.

(iii) Speakers preferred a larger object as the reference object for a smaller object. In describing the left arrangement, a substantial minority of subjects said *Behind the chair is a lamp*, whereas almost nobody said *Behind the lamp is a chair* in describing the right arrangement.

(iv) Speakers generally preferred to express relative location in terms of *in front of* rather than in terms of *behind*. Being IN FRONT OF is presumably a more salient relation than being BEHIND.

These findings show that there are preferred ways of expressing spatial relations propositionally, depending on intrinsic functional and gestalt properties of the scene. Moreover, Bürkle et al. (1986), using similar arrangements, showed that the place and the role of the interlocutor could affect the

perspective chosen. But all of these are tendencies, not iron laws. A speaker is free to choose one perspective rather than another. And indeed, the ways in which the same scene is described by different subjects are surprisingly variant. When one looks over Ehrich's protocols, one is struck by the fact that no two descriptions of the same furniture arrangement are identical. Each subject added a personal touch in terms of the objects, the relations, the qualities attended to, and the choice of perspective.

### 4.5.5 Acknowledging Language-Specific Requirements

Section 3.6 described some cases where there is obligatory grammatical encoding of particular conceptual features, even if these are irrelevant for communication. In a language that has a tense system, for instance, it is obligatory to encode in the preverbal message the deictic and intrinsic temporal properties of a state or event. This may not be part of the speaker's macroplanning, because it need not proceed from elaborating the communicative intention. Rather, in finalizing the preverbal message for expression the speaker will retrieve these pieces of conceptual information if they are not yet represented. However, this is necessary only for speech-act intentions of the types DECL(P) and ?(P); imperatives have fixed tense in English. The speaker will then, for declaratives and interrogatives, insert the two tense functions as proposition modifiers (see subsection 3.2.6).

We must assume that the speaker has at his disposal a set of routine procedures that perform this acknowledgment function automatically for whatever the language requires. It is unlikely that these computations require special attentional effort. And this may be a more general property of microplanning. Most of a speaker's attention is spent on macroplanning, on elaborating the illocutionary intention, and on retrieving information for expression.

The end result of microplanning is a preverbal message. There is no reason to assume that preverbal messages are delivered as integral wholes. In incremental speech production, bits and pieces of the message under construction may become available one after the other. Each bit is immediately picked up by the Formulator for grammatical encoding. As we will see in chapter 7, the order in which the chunks are delivered will affect the course of grammatical encoding.

## Summary

This chapter began with a characterization of communicative and illocutionary intentions. Performing a speech act involves more than just trans-

mitting information to an addressee; it also involves making a communicative intention recognizable from what is said. The process of generating messages was then analyzed as a two-step process, with macroplanning followed by microplanning. Macroplanning consists of elaborating the communicative intention in a sequence of goals and subgoals and, for each subgoal, selecting information whose expression will be a realization of the subgoal. Microplanning finalizes each speech act for expression by providing the message with an information structure that will guide the addressee in inferring the communicative intention.

Both macroplanning and microplanning are highly context-dependent, and the context of discourse is in continuous flux. We examined several aspects of the discourse context which speakers must keep track of in order to make relevant contributions. They must register the type of discourse they are involved in (such as an interview, or a lecture). They must also take into account the topic(s) of the discourse. Speakers also build an internal representation of the contents that have been contributed by themselves and by the interlocutors; this shared knowledge is called the speaker's discourse model. They also monitor what is currently in focus. And there is, finally, some record of what was literally said, especially where this was pragmatically important or salient.

The chapter then turned to macroplanning. It was discussed that selecting information for expression may involve substantial planning, memory search, and inference-making. A speaker's attention will move back and forth between performing these activities and finalizing the messages for expression. This can lead to characteristic rhythms in monological speech, which make it likely that most attention is spent on macroplanning.

A speaker can often make complex intentions recognizable by expressing rather limited amounts of information. The information must only be instrumental in inviting the intended inference on the part of the listener. We reviewed some experimental work on what speakers select for expression when they make reference to objects and when they construct requests. There turns out to be more redundancy in object naming than one would expect from Grice's maxims. But these deviations are not really irrational or unmotivated; overspecification of an object is helpful, especially when it reveals the *kind* of object focused by the speaker. Also, speakers give seemingly redundant information when they contrast the intended referent with a previously mentioned object. This is an effective way of guiding the addressee's search. The request studies also showed that speakers select efficient information for expression—information that acknowl-

edges the addressee's position in terms of willingness or ability to comply with the request and which, at the same time, makes it likely that the addressee will draw the intended inference. The elaboration of the original communicative intention may lead the speaker to conceive of what we called "side intentions," which may or may not be communicative in the restricted sense of this chapter. Side intentions tend to be encoded as rather independent messages, which can often be recognized as such in the speaker's discourse.

When the information to be expressed is complex, involving several successive speech acts, the speaker will have to decide on how to order the information for expression. This was called the speaker's linearization problem. Its solution depends, in the first place, on the content of what is to be expressed. A principle of natural order dictates default solutions for particular domains of discourse. The major example is that in the temporal domain events should be expressed in chronological order. In the second place, there are general restrictions on working memory that induce a speaker to prefer one linearization over another. These restrictions are well defined, and are quite general in nature.

Microplanning was the topic of the chapter's final section. The first aspect discussed was the assignment of an accessibility index to each referent in the message. This is done to inform the listener where the referent can be found: in the current focus, in the discourse model, somewhere else, or nowhere. Languages have a range of means for the grammatical encoding of this accessibility index. Also, a referent may be given the special status of topic. This is done to invite the interlocutor to store the new information under that referent's address in the discourse model. Referents may have varying degrees of saliency. A salient entity will have a better chance of getting an address in the discourse model than a nonsalient one. As a consequence, salient items are more easily topicalized than nonsalient ones. Items in the message will also vary in prominence. A speaker will make an item prominent if the listener has to store it as something new—a new referent, a new entity in a focused role, or a new predication.

Information selected for expression must eventually be encoded in a propositional format. This necessarily involves the assignment of perspective, in particular the choice of relations and reference points for these relations. These choices depend on various factors, such as the speaker's linearization strategy, the gestalt relations in a spatial scene, and the speaker's cognitive style.

Finally, the speaker will automatically retrieve the conceptual information to be acknowledged for the specific language spoken, such as temporal information when the language has a tense system. The final result of microplanning is a preverbal message that can be recognized by the Formulator as its characteristic input.

# Chapter 5
# Surface Structure

Once a message, or a fragment of a message, has been prepared for expression, the process of formulating can be initiated. Successive message fragments will trigger the Formulator to access lemmas, to inspect the message for functions, arguments, and modifiers, to specify grammatical relations, and to map these onto inflectional and phrasal structure.

This first stage of the formulating process was called "grammatical encoding" in chapter 1, and was distinguished from a second, phonological encoding stage in which word forms are accessed and prosodic patterns are generated. The present chapter will characterize the type of representation that forms the hinge between these two stages. It will be called "surface structure" (which involves an allusion but not a full commitment to particular grammatical theories). A surface structure is, by definition here, the output of grammatical encoding, and the input to phonological encoding. We will, however, stay rather close to Bresnan's (1982) notion of surface structure.

In order to understand the processes of grammatical encoding, which are discussed in subsequent chapters, we must have a sufficiently explicit specification of their target structures. It is, on the one hand, necessary to consider the way in which a surface structure expresses semantic relations through grammatical functions. This semantics-to-function mapping depends on the internal structure of lemmas, which are the terminal elements of a surface structure. It is, on the other hand, necessary to specify the way in which these grammatical functions are realized in a surface structure's hierarchical organization of phrases and in its case marking. This organization is essential input for phonological encoding.

These theoretical notions have, in general, not evolved from empirical analyses of the speaking process. They mostly stem from linguistics and computer science. Still, they do provide a much-needed framework for a theory of the speaker, which is not independently available. The theory we