Chapter 7

The Generation of Surface Structure

According to our lexical hypothesis, grammatical encoding is lexically driven. Conceptually activated lemmas instigate a variety of syntactic procedures to construct their proper syntactic environments. The procedures build the phrasal, clausal, and sentential patterns that express the required grammatical functions by means of phrasal ordering and case marking. Grammatical encoding is the process by which a message is mapped onto a surface structure.

In the present chapter we will first consider the kind of architecture needed for a Grammatical Encoder that is not only lexically driven but also incremental in its operation. The requirement of incrementality, as we saw in chapter 1, implies that a surface structure is, by and large, generated "from left to right" as successive fragments of the message become available. Wundt's principle requires that the generation of surface structure occur without much lookahead or backtracking, so that each surface unit produced can immediately be processed by the Phonological Encoder. The only explicit or computational theory of grammatical encoding that is both lexically driven and incremental in its operation is the Incremental Production Grammar (Kempen and Hoenkamp 1982, 1987; Kempen 1987; De Smedt and Kempen 1987; Hoenkamp 1983; for a review of other computational models of language generation, see Kempen 1988). The sketch of the Grammatical Encoder in section 7.1 will follow the main lines of this theory, which is a natural companion to Bresnan's Lexical Functional Grammar. The first aim of this sketch is to show that an incremental lexically driven Grammatical Encoder is a coherent and possible notion. The second aim is to demonstrate the notion's empirical potential. I will do this by mixing the presentation of the architecture with analyses of speech errors. Errors of grammatical encoding can be quite revealing of the underlying mechanisms. Are they consistent with the proposed architecture, or are they occasionally problematic?

After the rather "algorithmic" section 7.1, we will turn to the more classical research tradition in the psycholinguistics of grammatical encoding. Section 7.2 will deal with units of grammatical encoding. What do pauses and speech-onset latencies tell us about "chunks" of grammatical encoding? Section 7.3 addresses the question of how the speaker assigns grammatical functions to more or less prominent arguments in the message. The accessibility of concepts and the need to topicalize can affect the course and the outcome of grammatical encoding. Section 7.4 reviews some research on cohesion in grammatical encoding. How does the speaker make use of referential expressions and syntactic means in order to make his ongoing sentence cohesive with previous discourse? Section 7.5 approaches the issues raised in chapter 1: whether grammatical encoding can feed back to message encoding (this should not be the case in a modular theory) and whether grammatical encoding can, in turn, be affected by feedback from the next processing component—phonological encoding.

For a solid review of the psycholinguistic research in grammatical encoding, see Bock 1987a.

7.1 The Architecture of Grammatical Encoding

7.1.1 Some Basic Kinds of Operation

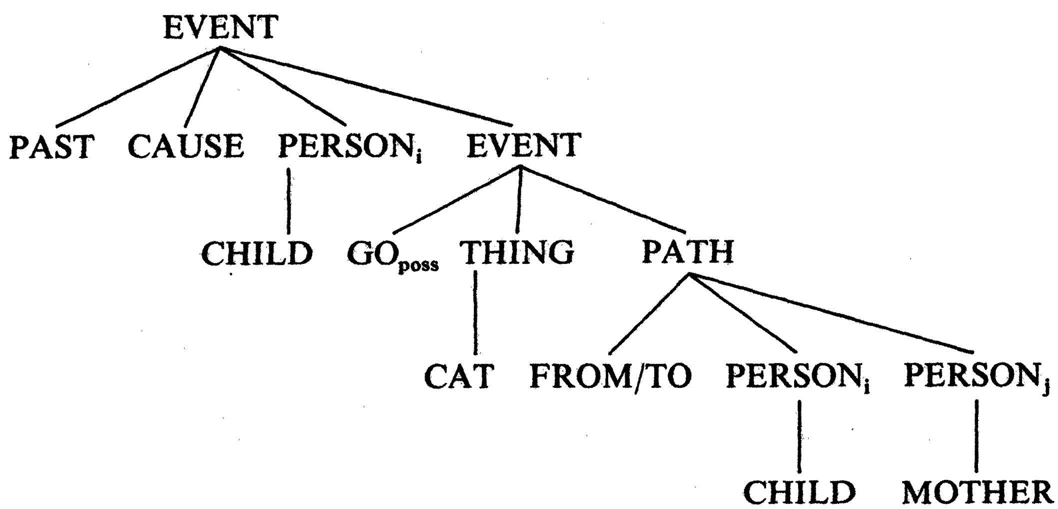
The claim that grammatical encoding is lexically driven implies that the encoding operations are largely controlled by the grammatical properties of the lemmas retrieved. It does not mean that lexical elements are *procedures*. The previous chapter treated lemmas rather as bundles of *declarative* knowledge about a word's meaning and grammar. This declarative knowledge becomes available when a lemma is retrieved. The grammatical encoding procedures of the Formulator will then be guided by the information the lemmas make available. The lemma *give*, for instance, requires slots for its subject and for its direct and indirect objects. This causes dedicated syntactic procedures to set up the appropriate frame. More generally, lemmas call specialized syntactic procedures in some orderly fashion, so as to produce a unified surface structure as eventual output. It is in this vein that Kempen and Hoenkamp (1987) made a procedural package of what verbs require of their syntactic environment, and similar packages for nouns, adjectives, and prepositions. These packages are the specialists that build S-, NP-, AP-, and PP-constituents around the major category lemmas. They contain the necessary procedural knowledge for dealing with the grammatical constraints that the different kinds of lemma impose on their syntactic environment.

Our initial sketch of the architecture proposed by Kempen and Hoenkamp will be made from a concrete example: the generation of the simple declarative sentence *the child gave the mother the CAT* (with pitch accent on CAT), for which the groundwork has already been laid in chapters 5 and 6. Since the example is an English sentence, some adaptations of the algorithm will be necessary—in particular, the addition of a VP-procedure (which is probably not necessary for the generation of Dutch, Kempen and Hoenkamp's target language). There are seven major kinds of procedure involved in the generation of such a sentence.

1. A lemma is retrieved if a fragment of the message satisfies its conceptual specifications. This semantically conditioned retrieval of lemmas was discussed extensively in chapter 6.

The message underlying the example sentence was given in figure 5.1a and is repeated here.

message:



In chapter 5 the accessibility status of the referents in this message was ignored. Here we will assume that all three referents (CHILD, CAT, and MOTHER) are "+ accessible." We will, moreover, assume that CAT is "+ prominent" (see subsection 4.5.1). Remember, moreover, that the left-to-right ordering is merely a notational convention. The order in which the message fragments become available may be a different one. That order can be a major determinant of the eventual syntactic form, as we will see.

Suppose that the concept CHILD was the first message fragment delivered by the conceptualizer. Only the lemma *child* out of the speaker's 30,000 or so lemmas finds its conceptual conditions satisfied. It is retrieved, and its internal grammatical specifications become available.

2. The lemma's syntactic category calls a *categorial procedure*, namely a building instruction for the phrasal category in which the lemma can fulfill the function of head. This holds for lemmas of types N, V, A, and P; the categorial procedures are NP, VP, AP, and PP, respectively.

The lemma *child* is of syntactic category N, and N can only be head of a noun phrase, NP (see subsection 5.1.3). So, *child* calls procedure NP, which is specialized in building noun phrases.

Categorial procedures can be called not only by lemmas but also by higher-order functional procedures—procedures that try to realize a certain head-of-phrase function (e.g., predicate for S, main verb for VP, head noun for NP) or a certain grammatical function (e.g., subject, direct object).

3. The categorial procedure inspects the message for conceptual material that can fill its obligatory or optional complements and specifiers and can provide values for diacritic parameters. Categorial procedures can also inspect the message for relevant head-of-phrase information.

The NP categorial procedure with *child* as head looks for modifying or specifying information attached to the concept CHILD in the message. In particular, it checks the accessibility status of CHILD, and it finds the value "+ accessible". In addition, since *child* is a count noun, the procedure inspects the concept for number and comes up with the parameter "singular". The NP procedure derives the parameter value "third person" from the lemma's syntactic category N.

4. The categorial procedure transfers control to subroutines—functional procedures for handling all the complements, specifiers, and parameter values it has found. These subroutines work in parallel.

In the example, NP calls upon the functional procedure DET, which can generate a definite specifier expressing the conceptual information "+ accessible". DET calls the lemma *the*, which consists largely of a pointer to the word form. In other languages, such as French and Dutch, DET will have to inspect the gender and number parameters of the NP-head and insert their values in the list of diacritic parameters of the article lemma. This is, of course, required for finding the appropriate word form (*le*, *la*, or *les* in French; *de* or *het* in Dutch). In German and various other languages, the functional procedure DET also needs case information, since the word form of the article depends on the grammatical function of the NP by which it was called (subject of the sentence, object of the prepositional phrase, etc.). So far, this grammatical function is still undetermined for *the child.*

NP also hands the "singular" information to the head-of-phrase procedure, which inserts it as a diacritic parameter value in the lemma for *child.*

Generally, the functional procedures deliver their results to the categorial specialist that called them. In the example, DET delivers the lemma *the* to the NP, and the head noun procedure does the same with *child.*

5. The categorial procedure determines the order of the materials it receives back from the functional procedures it called. Each categorial procedure has a "holder" with a number of slots, and there are certain restrictions on the order in which the slots can be filled by the procedure. There is, however, an important psychological principle here: *Materials will be put in the leftmost slot they are allowed to occupy.*

The latter is an implementation of the "left-to-right" incremental principle. Fragments of a message are grammatically expressed as much as possible in the order in which they became available. As we will see in the more empirical sections of this chapter, topical or salient concepts tend to be expressed early in the sentence. In terms of the model, one could say that their categorial procedures are initiated early because their lemmas are available early. Left-to-right filling of contiguous slots in the holders of categorial procedures at the earliest possible moment contributes to the fluency of speech, because phonological encoding can proceed "on line" (i.e., incrementally from left to right).

In the example, the NP procedure puts the lemma *the* to the left of the lemma *child.* This is an obligatory order for English. However, the fact that the whole NP *the child* will end up as the leftmost constituent of the sentence is due not to an obligatory ordering but to Wundt's principle (subsection 1.5.2). CHILD being the first conceptual fragment to become available causes the noun phrase *the child* to end up in sentence-initial position.

6. The categorial procedure chooses a grammatical function for its output; that is, it decides on its *functional destination.* It will become a head or a complement of some higher-order categorial procedure. Higher-order categorial procedures are ones that do not have lemmas but phrases as heads. They are NP′, VP′, AP′, PP′, and S (the latter takes a VP or VP′ as head or predicate). The main rule for functional destination is that it is precisely the head or complement function of the higher-order procedure that called the categorial procedure to start with. If there is no such higher-order calling procedure, there is a preferential destination (and in fact a preferential destination *order*) for each categorial procedure.

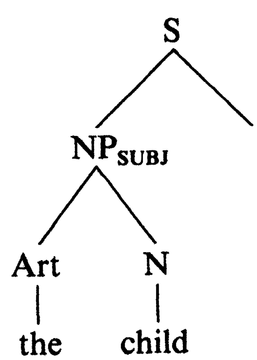
Let us see how this works in the above example. Remember that the NP categorial procedure that generated *the child* was called by the noun lemma *child*, not by any higher-order categorial procedure. The default destination for the output of the NP procedure is "subject of S". Or, in the format of chapter 5: The NP acquires the subscript SUBJ and hence subordinates itself to the categorial procedure S. This provides NP, at the same time, with the diacritic case feature "nominative", which is then available for the procedures DET and N when their lemmas require diacritic case values. (They don't in the present example.) Later we will consider cases where there *is* a higher-order calling procedure.

7. The preferential higher-order categorial procedure is activated by receiving the output of a lower-order procedure.

In this way, NP_{SUBJ} calls the categorial procedure S, which is a specialist in building sentences. S accepts the noun phrase as its subject, and registers its person (third) and number (singular) parameters. S also inspects the message for a mood marker. It finds none, which induces S to impose the default declarative constituent order (see subsection 3.5.1). S can then assign the subject noun phrase to its leftmost slot by an operation of type 5. (When there is a mood marker, IMP or ?, this will lead to a diacritic feature on the verb. Such a mood-marked verb calls the appropriate categorial procedure for imperative or interrogative word order.)

These are the seven main kinds of procedures underlying the architecture of grammatical encoding in the Kempen-Hoenkamp model. Each individual procedure can be written as a production—a condition/action pair of the kind IF X THEN Y, where X is the condition and Y is the action.

Let us see what stage we have reached in the generation of the surface structure of *the child gave the mother the CAT*. It is easily displayed by the following bit of tree structure:



This may not seem much; however, *the child* has reached its final, leftmost position in the sentence. There is nothing to prevent its phonological encoding while the grammatical encoding for the remainder of the sentence is still in (parallel) operation. This is incremental production.

The following encoding steps are essentially of the same kind as the previous ones. The categorial procedure S, once called, will execute procedures of type 3. It will, in particular, scan the message for a conceptual function that has CHILD as an argument. The only function in the message satisfying this property is CAUSE. More precise:

CAUSE (PERSON_i, EVENT)

where PERSON_i = CHILD.

The procedure S also inspects the modifiers of the CAUSE function to set its tense parameter (and similarly for aspect, which will be ignored here). It finds PAST (which was shorthand for deictic $r < u$ and intrinsic $e = r$; see subsections 2.2.3 and 3.2.6). This induces S to assign the value "past tense" to its tense parameter.

Having localized the function of which CHILD is an argument and registered its temporal modifiers, S turns to a type 4 operation. It calls the functional procedure that can construct its head of phrase, or predicate: the PRED procedure. S provides it with the collected diacritic parameters: "past tense", "third person", and "plural". S reserves its second slot for the predicate output: the verb phrase (the first slot being occupied by the subject NP). Since this slot is consecutive to the already-filled leftmost slot, S allows the PRED procedure to make its output incrementally available to the Phonological Encoder.

PRED, in its turn, calls the categorial specialist VP, which inspects the message to localize materials for its head of phrase, the verb. The structure of the tree emanating from the CAUSE function is compatible with the conceptual specifications of *give* (see subsection 6.2.1). But it is also compatible with the semantics of *receive* (subsection 6.2.2). However, only a lemma that maps the agent onto SUBJ of S can comply with the already-generated functional destination of the NP *the child.* The lemma *receive* would assign the SUBJ function to the goal argument, which is inappropriate. The lemma *give*, however, makes the mapping

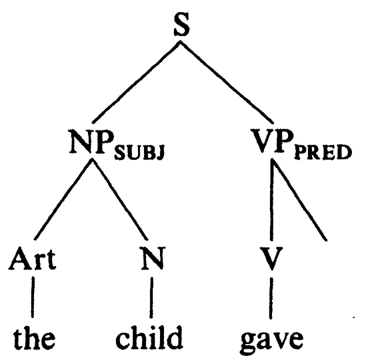
```
X(agent), Y(theme), Z(goal)
  |          |         |
SUBJ        DO        IO
```

(and we momentarily disregard the other lemma for *give*, which also assigns the agent a subject role; see subsection 6.2.2). VP accepts this lemma as its head of phrase.

Procedure VP transmits its diacritic parameters to V, its head of phrase, which initiates the functional main-verb procedure. It assigns the parameters to the main-verb lemma *give*. The parameters are set to create a so-called tensed verb, i.e., a verb which is marked for tense and which agrees with the sentence subject in terms of person and number. This being completed, the main-verb procedure will report its output to VP, which, by an operation of type 5, assigns the first or leftmost slot in its holder to this lemma, by which it becomes available for phonological encoding.

In the example we are constructing, the retrieval of *give* was initiated by a syntactic procedure (the VP procedure). There was a need for a main verb. But *give* could also have been retrieved without syntactic intervention, in just the same way as *child* was—that is, by mere "resonance" with the corresponding fragment of the message. In that case, the retrieved verb lemma would have called the categorial procedure VP, which would eventually have delivered its output to S, its preferential destination.

How far have we proceeded? After the lemmas *the* and *child*, *give* has now reached its final destination in the surface structure. The situation is this:



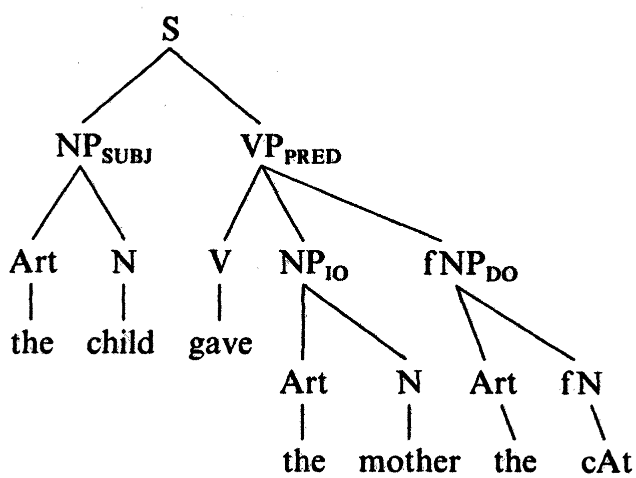
The phonological properties of *give* with parameters "3rd person", "singular", and "past" (the graph presents this as *gave*) can now be retrieved by the phonological encoding procedures concurrent with the further generation of the surface structure.

Next VP calls the functional procedures direct object (DO) and indirect object (IO) to inspect the message (operations of type 3). This is done in order to find the *theme* and the *goal*, respectively, as required by the head verb. Remember that the lemma for *give* dictates the *theme* to map onto a direct-object complement, and the *goal* onto an indirect-object complement.

The functional procedures DO and IO, which may run in parallel, identify CAT and MOTHER (respectively) in the message. They both call the categorial procedure NP to lexicalize these conceptual fragments. The NP procedure is of type 3, and it may run in parallel for DO and IO. Both running NP procedures call the head-noun procedure to inspect the message fragments made available by DO and IO. This can, again, be done in parallel fashion. It leads to the retrieval of the lemmas *cat* and *mother*, respectively. Meanwhile, the NP-procedures perform operations of type 3; i.e., they look for modifying or specifying information attached to the message fragments they are working on (CAT and MOTHER). Both register the property "+ accessible", and the one working on CAT also finds the feature "+ prominent". The former feature leads to the NP's calling a definite DET procedure. The registered prominence of CAT causes NP to produce a focus feature (f). The head-noun procedures deliver their lemmas (*cat* and *mother*) to their respective NP procedures, which accept these lemmas as heads of phrase. From here on, both NPs are constructed in the way we have seen for the NP *the child.* The head of phrase *cat* receives the focus feature (f), which will, in turn, be translated into a pitch-accent parameter for the lemma *cat* in accordance with the rules discussed in section 5.2. The parallel-running NP procedures have DO and IO as functional destinations for their output in the type 6 operation. This is because they had been called to fill these VP slots to start with.

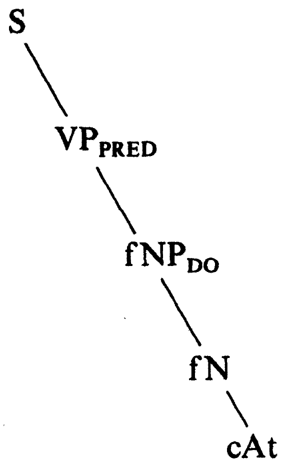
When the functional procedures DO and IO receive the trees the NP procedures have built (i.e., those for *the cat* and *the mother*), they assign accusative and dative case, respectively, just as SUBJ assigned nominative case to the NP for *the child.* It is important to note that, according to this algorithm, case is assigned at a fairly late stage—namely *after* the NPs return their trees to the functional procedures that called them. One motive for organizing the control structure in this way derives from the nature of certain speech errors; we will return to this shortly.

Operations of type 5 make VP arrange the NPs in the order IO → DO. Finally, VP turns to a type 6 procedure, choosing the predicate functional procedure that called it as its functional destination. The predicate information is then delivered to S, which puts it in the second slot of its holder. Because nothing remains to be covered in the message, S halts the procedure. The produced surface structure is



(where *gave* is shorthand for *give* with its diacritic parameters, and *cAt* stands for *cat* plus pitch accent).

This picture is misleading insofar as it is a static representation of what was built up and transmitted over time. If one were to draw only the *active* nodes at any one moment, one would never see the whole graph. At the final stage, just after NP_{DO} has put the lemma *cat* in its second slot, the only active part of the tree is



The Kempen-Hoenkamp algorithm shows a strict alternation in the calling of procedures: Categorial procedures call functional procedures, and functional procedures call categorial procedures. Also, a categorial procedure always delivers its output to the functional procedure that called it (if any), and normally a functional procedure delivers its output to the categorial procedure that called it.

Lemma structure plays a central role in the generation of surface structure. In particular, the main verb dictates what arguments have to be checked in the message, and which grammatical functions will be assigned to them. In the Kempen-Hoenkamp algorithm the activation of lemmas is, to some degree, guided by procedures of type 3. The control structure of the encoding operations dictates which parts of the message are accessed for gramatical encoding at what moments. This may activate certain lemmas that would otherwise have stayed asleep for a while.

But the control structure itself depends on the availability of the different message fragments (incremental production). In the example, CHILD was available early in the message. It activated the lemma *child* without the intervention of a type 3 procedure. As a consequence, the first categorial procedure to run was NP. The order in which grammatical encoding procedures operate depends, in part, on the order in which lemmas become available. And this, in turn, depends on the order in which message fragments are produced by the speaker. Fluent speech requires incremental production. Lemmas are grammatically and phonologically processed as soon as possible after retrieval. Subsequent grammatical encoding procedures will adapt to already available structure rather than revise it. An example is the choice of *give* after the NP *the child* had been generated. To keep the NP in its leftmost position, the lemma *receive*—which did fit the conceptual specification in the message, and might have been used, was filtered out by the categorial procedures. Its use would have created extra waiting time in order to construct the NP *the mother* for appearance in leftmost position, and for revising the subject assignment of *the child.* It is probably too strong to say that such "on the fly" revisions never occur; they are, however, kept to a minimum.

Incremental production can also explain the choice of a passive lemma for *give* when *cat* would have been the first available lemma. The noun *cat* would have called NP, and this categorial procedure would have claimed the subject function for *the cat*. S would have assigned this NP its leftmost slot. But then the passive-verb lemma, which maps the *theme* argument onto SUBJ (subsection 6.2.2), would be the only appropriate choice for further processing. The final result would have been the surface structure corresponding to *The cat was given to the mother by the child.* It is, of course, an important question what determines the availability or primacy of a lemma. It clearly has to do with the topichood of the corresponding concept in the message. This factor and other determinants will be considered in section 7.3.

Parallel processing is a main contributor to fluency. In the example, the direct-object and indirect-object constituents were constructed in parallel. There are no control problems here as long as the active procedures are all connected, as in the example. But imagine the situation in which both CHILD and CAT are available early in the message such that both lemmas *child* and *cat* are independently and more or less simultaneously retrieved. Both will call NP procedures. These, in turn, will both deliver their output to S as a preferential destination; both will claim the subject function for their output (*the child* and *the cat*, respectively). The simplest control measure is that prior entry will decide such cases. If the NP procedure that constructs *the child* is the first to call S, the other one (constructing *the cat*) will be given its second preferential destination. This is a procedure of type 6. This second preference will be the direct-object function. The only verb lemma that can satisfy both the conceptual specifications in the message and the grammatical functions now assigned to the conceptual arguments is the alternative active lemma for *give* (which has so far been ignored in the discussion). It will then govern the generation of the sentence *The child gave the cat to the mother*, where *cat* is the direct object. Empirical evidence in support of this analysis will be presented in section 7.3.

It is not obvious that the rate of left-to-right delivery of surface structure fragments corresponds perfectly to the rate of the subsequent phonological encoding of these fragments. Hence, fragments may occasionally have to wait for further processing. This requires a buffering facility for output of grammatical encoding: the Syntactic Buffer. The Phonological Encoder retrieves subsequent fragments from this buffer.

7.1.2 Speech Errors: Exchanges of Same-Category Phrases and Words

The algorithm so far can provide a better understanding of how certain word and phrase exchanges arise in fluent speech. In chapter 6 word exchanges were adduced to parallel processing of different message fragments. In the example above, there is parallelness in the grammatical encoding of MOTHER and CAT. The two NPs are constructed simultaneously, according to the algorithm. Let us consider how the following error could arise:

(1) The child gave the cat the mOther [*instead of*: The child gave the mother the cAt]

The error may arise at an early stage of generation, namely when the two NP procedures accept their head nouns. We saw that, according to the algorithm, each NP called a head-noun procedure that would retrieve the lemmas for the corresponding arguments in the message (MOTHER and CAT). These head-noun procedures may err in the destination of their output, the lemmas *mother* and *cat*. The NP procedures blindly accept whatever noun is made available at the time they need one. In a parallel-processing environment, this may on occasion be the wrong one. The only thing that counts at this level of processing is that the head slot of NP is filled by a lemma of syntactic category N.

The consequences of such a destination error are quite interesting. First, the NP working on CAT will equally blindly assign its focus feature (f) to its head noun, which is now the lemma *mother*. As a consequence, *mother* will receive pitch accent. This is called *stranding*: The pitch accent stays in place, detached from its exchanging target lemma. The replacing element (*mother*) is said to *accommodate* the stranded feature. Second, each NP will call the appropriate DET procedure, i.e., the procedure involved in expressing the intended referent's accessibility status. The effect is invisible in example 1, because both NPs required definite determiners. But if the target sentence had been *The child gave the mother a cAt*, the exchange error would have produced *The child gave the cat a mOther.*

And indeed, real word reversals tend to show both properties . Examples 2 (from Fromkin 1973) is a case in point.

(2) Seymour sliced the knife with a salAmi [*instead of*: the salami with a knIfe]

Here the nouns *knife* and *salami* are exchanged. Stranding of accent occurs, according to expectation. In fact, Fromkin noted that this is the general rule in word reversals. The accent-stranding phenomenon—probably first noticed by Boomer and Laver (1968)—has been extensively analyzed by Fromkin (1971), Garrett (1982),* and others. In addition, the example error shows stranding of (in)definiteness: The determiners *the* and *a* appear in the right places, but with the wrong nouns.

What should happen to the number feature, on the present theoretical analysis? It was suggested that, by a procedure of type 3, the NP procedure registers the conceptual information that would be relevant for computing the number parameter of its head count noun. And, by a procedure of type 4, NP hands the computed number feature to the head-of-phrase procedure, which inserts it as a diacritic parameter in its lemma. This predicts that number should strand in word reversals as well. When NP accepts the wrong head of phrase, it will blindly provide it with the originally derived number feature. And indeed, such cases occur. Here is an example from Fromkin 1973:

(3) a hole full of floors [*instead of*: a floor full of holes]

If the number feature had exchanged as well, the result would have been *holes full of a floor*.

However, there are cases where number does move with its "own" lemma, as in the following (from Fromkin 1971):

(4) examine the horse of the eyes [*instead of*: the eyes of the horse]

The error was not *examine the horses of the eye*. Stemberger (1985a) reported that stranding of the plural, as in example 3, is about four times as frequent as non-stranding cases, such as example 4. This suggests that examples 3 and 4 have different etiologies. In example 4 we probably have to deal with an exchange of whole noun phrases (*the horse* and *the eyes*) instead of a mere exchange of head nouns. Where could this arise in the

*Garrett (1982) gives examples that suggest that there is less or no stranding of *contrastive* accent. This agrees with Cutler and Isard's (1980) special treatment of contrastive stress, referred to in subsection 5.2.2.

sketched algorithm? It could occur at the later stage where the NP procedures deliver their output to their functional destinations, direct object (of *examine*) and oblique object (of *of*). These are procedures of type 6. If the DO and OBL (oblique object) procedures are, in parallel, waiting to receive a noun phrase, a destination error may occur: DO accepts *the horse* and OBL takes *the eyes*, probably because of some slight mistiming.

That destination errors for complete NPs indeed occur is apparent from the following two examples:

(5) I have to smoke my coffee with a cigarette [*instead of*: I have to smoke a cigarette with my coffee] (from Fromkin 1973)

(6) I got into this guy with a discussion [*instead of*: I got into a discussion with this guy] (from Garrett 1980a)

In each of these cases there is an exchange of full noun phrases: *my coffee* and *a cigarette* in example 5, *this guy* and *a discussion* in example 6.

If indeed the noun phrases were delivered to the wrong functional destinations, one would expect a stranding of case marking but not a stranding of plural. Remember that the algorithm assigns case to a constituent only *after* it has been returned to the functional procedure that instigated its construction. The speech-error data are not conclusive in this respect. English has rather poor case morphology, and the published speech-error corpora do not contain much relevant material. One interesting example can be found in Fay (1980b):

(7) If I was done to that [*instead of*: If that was done to me]

Here the subject (*that*) and the prepositional object (*me*) switched places. But the error did not become *If me was done to that*; the first person pronoun was given nominative case (*I*), as it should in a subject function. This is in agreement with the algorithm.

German is richer in case marking, and indeed Meringer and Mayer's (1896) corpus contains a few word exchanges involving case. Some of them clearly follow the predicted pattern of case stranding, such as the following:

(8) Bis er es bei Dir abholt [*instead of*: Bis Du es bei ihm abholst] (*English translation*: Until he collects it from you)

If case had moved with the pronouns, the speaker would have said *Bis ihm es bei Du abholt*. However, the subject pronoun receives nominal case marking, as it should. Similarly, the prepositional object (*Dir*) has the correct dative case. A more recent, negative example is this, from the Bierwisch corpus:

(9) Wenn der Wand aus der Nagel fällt [*instead of*: Wenn der Nagel aus der Wand fällt] (*English translation*: when the (dative) wall falls from the (nominative) nail)

If case had stranded, the error would have been *Wenn die Wand aus dem Nagel fällt* (*When the* (nominative) *wall falls from the* (dative) *nail*).

Berg (1987) reports other examples of failing case accommodation in German. Where such accommodation does not occur, one must conclude that case is assigned *before* the noun phrases are delivered to their functional destination.

Generally, however, case, number, and other inflections have a strong tendency to be stranded (Garrett 1982, Stemberger 1983a, 1985a). This is supportive of the notion that the corresponding diacritical features are assigned *after* the lemma is inserted into its grammatical slot.

Berg (1987), in a careful analysis of accommodation in various languages, concluded that, as a rule, an erroneous element adapts to its new environment, as Garrett (1980b) suggested. But the environment usually does not adapt to the erroneous element. This is precisely the pattern of results one would expect if diacritic features are assigned to the lemma "from the environment." In example 3 above, for instance, it is the NP procedure that assigns the plurality feature to its head noun, whatever noun lemma happens to show up. Examples 4 and 5 do not contradict this, because here whole NPs—not just nouns—are delivered to the wrong destination.

Analyses of this kind show the advantage of an explicit algorithm for grammatical encoding: It helps us predict particular *kinds* of speech error, and to exclude others as impossible. It can even tell us something about the mutual dependency of different kinds of speech error, and therefore on their relative frequencies. For instance, on the above analysis, if plural strands, then case should strand as well, but not vice versa. This is because the assignment of the plural feature occurs at a lower level of destination than the assignment of the case feature. Violation of such predictions will force us to abandon or revise the model.

The Kempen-Hoenkamp Incremental Production Grammar assumes the existence of "syntactic specialists," a collection of grammatical and functional procedures. These procedures, we saw, function in a highly modular fashion. They do their thing when called and provided with their characteristic input, and they deliver their own special kind of output to a particular address. They do their work automatically (i.e., without using the speaker's attentional resources) and in parallel fashion. And they are stupid. They blindly accept the wrong input if it is of the right kind. The speech errors discussed so far testify to this witless automaticity of grammatical encoding.

7.1.3 Some More Complex Cases

In Chapter 6, a few remarks were made on "raising" verbs, such as *believe*. It was observed that in a sentence like *Attila believed the world to be flat*, the verb in the main clause imposes certain restrictions on the structure of the complement: that the subject of V-COMP should be realized as the direct object of the main clause, and that the tense of V-COMP should be infinitive. Let us see what is needed to generate such a sentence, given the seven kinds of procedures discussed in subsection 7.1.1.

Much depends on which lemma is triggered first. Let us assume that *Attila* sets the stage. *Attila*, being a noun, calls a categorial procedure in which it can figure as head, according to a type 2 operation. This procedure is NP, which accepts it as a proper (head) noun. NP does not call DET in case of proper nouns, but immediately chooses a default destination—namely SUBJ (a type 6 operation). This activates the categorial procedure S (type 7), and S puts the acquired subject in the leftmost slot of its holder. S then scans the conceptual structure for a function of which ATILLA is an argument (type 3). It comes up with the concept BELIEVE, which has two conceptual arguments: an experiencer and a state of affairs (or *theme*). This fragment, plus the information that SUBJ = experiencer, plus the diacritic parameters for person, number, and tense, are then handed over by S to its head-of-phrase procedure (PRED), which calls VP.

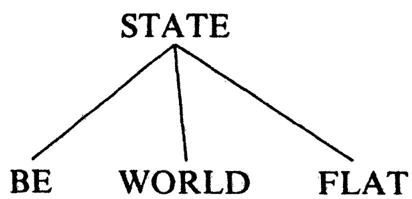
VP's main verb procedure causes BELIEVE to trigger the retrieval of the lemma *believe*. This lemma specifies three grammatical functions: SUBJ, DO, and V-COMP (see subsection 6.2.1). Can VP accept this lemma, in view of the present state of generation? Yes, since the only existing restriction comes from the already-available information on the subject as experiencer. The lemma specifies that the SUBJ function should be reserved for the conceptual experiencer argument, and that is exactly the case.

Given the lemma *believe*, VP must reserve two further slots in its holder. It reserves the second one for DO and the third one for V-COMP. But here something special happens. VP does not call the procedure DO, but only V-COMP (a call-procedure operation of type 4). It reserves its DO slot for V-COMP's subject; this is what the structure of the lemma dictates. For the same reason, the call to V-COMP specifies that it should be infinitival.

V-COMP calls categorial procedure S to inspect the state-of-affairs argument. The S procedure is also provided with the diacritic feature

"+ infinitive". This feature suffices for S to "know" that the functional destination of its subject noun phrase is DO of the main VP; there will be no SUBJ functional destination for that noun phrase. Also, that feature causes the functional destination of the predicate VP to be the V-COMP slot of the main VP, not the predicate slot of S. In other words, V-COMP's S procedure will behave in a deviant fashion. It will not set up a holder of its own, and it induces external destinations for its subject NP and predicate VP. We will soon see the consequences.

Inspecting the state-of-affairs argument of BELIEVE, S finds a STATE function like this:



S calls a SUBJ and a PRED functional procedure, and assigns the fragments WORLD and BE FLAT to them, respectively. The functions may or may not work in parallel. The SUBJ procedure calls the categorial procedure NP to lexicalize the WORLD fragment. This leads to the retrieval of *world*, and to the eventual construction of the noun-phrase tree *the world*, which is handed back to SUBJ. SUBJ, in turn, delivers it to the DO slot of the main VP, where it receives case. This case assignment is not quite visible in the present sentence, but an otherwise equivalent sentence shows this clearly: *Attila believed her to be flat*. Here *her* has the accusative case required for direct objects, not the nominative case of subjects (as in *Attila believed she to be flat*).

The PRED procedure, meanwhile, has called the categorial procedure VP to lexicalize its message fragment BE (FLAT). The main-verb procedure comes up with lemma *be*, which requires a subject. (If the SUBJ procedure had not started independently, it would have been called now.) The main-verb procedure not only realizes "infinitive" as a (zero) inflectional parameter on the lemma *be*; it also activates the preposition *to*. VP accepts this ordered pair and puts it in its leftmost slots.

The head verb *be*, finally, requires an adjectival complement, on which the property (FLAT) in the message is to be mapped. Hence, VP calls an AP procedure to lexicalize FLAT. This leads to the retrieval of the adjective *flat*, which becomes the head (and only element) of the adjective phrase, ending up in VP's third slot. VP now has its holders filled, and the PRED procedure that called it will deliver this output to the main-clause VP. It puts this verb phrase in the latter's third slot, which was reserved for the

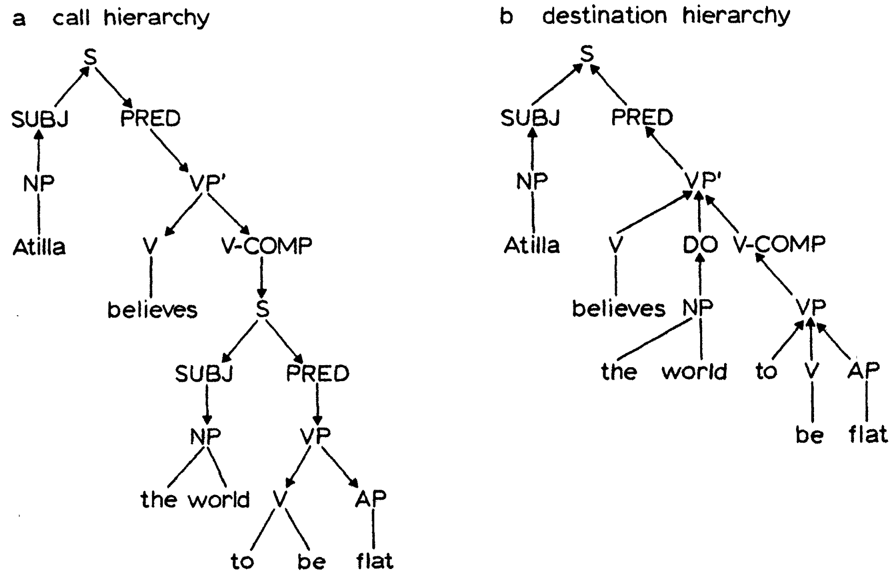


Figure 7.1
Call hierarchy and destination hierarchy for the sentence ***Attila believed the world to be flat***.

output of V-COMP. The output of the main-clause VP will, finally, end up in the main S-predicate slot, and S will halt.

It is useful to distinguish the two verb phrases generated in this example. The main verb phrase will be denoted by VP′ and the other by VP, in accordance with the notation proposed in subsection 5.1.3.

The above example is important because the *call hierarchy* differed from the *destination hierarchy*. This distinction (introduced in Hoenkamp 1983) is probably relevant for the interpretation of certain experimental findings in sentence production, as will be discussed in section 7.2. Figure 7.1 presents the two hierarchies for the present example sentence. In 7.1a the arrows indicate which procedures called which; in 7.1b the arrows indicate where the procedures delivered their output (certain less relevant details are left out).

The major difference between the two hierarchies is that the call hierarchy has an embedded call to S, but in the destination hierarchy nothing is delivered to S at that level. In other words, the embedded S procedure is involved in inspecting the message for subject and predicate information, but it is not involved in ordering the output of the procedures it called; the ordering of constituents is exclusively handled by VP′. Let us call these the *inspecting* and *ordering* tasks of the S procedure, respectively.

The difference is also reminiscent of the distinction made by rather diverse linguistic theories between *deep structure* and *surface structure*. The point will not be elaborated here; see Hoenkamp 1983 and Kempen and Hoenkamp 1987.

The distinction between call and destination hierarchies can be exploited with profit to handle other cases as well. The S-COMP of certain adjectives (as in *John is eager to please*) and the V-COMP structure of auxiliaries (as in *Olithia can go*), discussed in subsections 6.2.3 and 6.2.4, are handled in rather analogous ways. In questions (*Can Olithia go?*), S will order the output of VP's constituents (*can* and *go*). So-called *Wh-movement*, as in *Who do you think John saw?*, where the *who* in front is the output of the object procedure of the embedded sentence (and where an "empty element" is left after *saw*—see subsection 5.1.3), can also be handled in an elegant way; see Kempen and Hoenkamp 1987. It is essential to the character of all these solutions that the speaker can incrementally produce from left to right, without any backtracking.

7.1.4 Ellipsis

The algorithm can also account for the incremental production of ellipsis in coordinative sentences (e.g., *John met the president, and Peter the secretary*). It involves skipping certain otherwise iterative operations in the generation of the second clause. We will not pursue this issue here, but see De Smedt and Kempen 1987. This is also the main mechanism involved in the generation of repairs, to which we will return in chapter 12.

Somewhat less straightforward are the cases of ellipsis discussed in subsection 3.2.7. There we considered the possibility that messages can be of any semantic type—proposition, entity, predicate, or whatever. Take an entity message, e.g., CHILD. If it has the feature "+ accessible", the algorithm discussed will generate the noun phrase *the child*, and the NP procedure will deliver it as subject of S, by which S is called into operation. S gives the noun phrase nominative case and inspects the message for a function of which CHILD is an argument—in vain this time; there is nothing more to be found. It can, of course, be specified that the procedure will halt in such a state. But is this what we want? Let us return to the example in subsection 3.2.7 where the interlocutor asks *Who did Peter visit?*. If the answer message is the same (i.e., CHILD), then a seemingly correct surface structure is generated: *the child.* But it carries the wrong case. This would come to the surface in a pronominal answer, which could be *her* (accusative) but not *she* (nominative).

How could it be organized that the NP procedure claims the object function for *the child*? Would it be enough to mark CHILD in the message as patient? The algorithm should then be extended so as to claim object slots for patients in messages. But this is clearly not what we want either. A patient can easily appear in subject position (as in *The child was visited by Peter*), and the algorithm would no longer be able to generate such cases correctly. What is worse is that in certain languages the same thematic role will be expressed in different cases, depended on the verb used. A good German example involves the quasi-synonymous verbs *treffen* and *begegnen* (*to meet*):

Question	Answer
Wirst Du Helmut treffen?	Nein, den Präsidenten (accusative)
Wirst Du Helmut begegnen?	Nein, dem Präsidenten (dative)
(Will you meet Helmut?)	(No, the president)

The elliptical answers have the case marking that is required by the verb. This forces us to recognize that the elliptical production does, in some way or another, involve the particular verb lemma, in spite of the fact that it is not overtly used in the answer. A categorial procedure can apparently inspect the surface structure of the question to which it is generating an elliptical answer. There are no concrete proposals as to how this is done.

7.1.5 Ordering Errors

The categorial procedures S, NP, VP, AP, and PP have holders containing more than a single slot. The procedures are specialists in ordering returned constituents in these slots. The control structure of incremental production guarantees that the ordering is, as much as possible, done from left to right as materials are returned, but there are obligatory restrictions which may lead to necessary inversions. I have not discussed in any detail the ordering rules applied by the different categorial procedures, but it should be noted that the categorial procedures do not have a single fixed ordering scheme. This is especially apparent for S, which handles questions and declaratives and main and subordinate clauses in different ways. Given the existence of such closely related but different ordering operations in the algorithm, one might expect a vulnerability to specific ordering errors in speakers, namely the imposition of a possible though inappropriate order. Speakers do make ordering errors that are interpretable from this perspective. Fay (1980b) collected and analyzed several cases of this sort. Examples 10 and 11 are from his collection.

(10) I wonder how can she tell. [I wonder how she can tell.]

(11) Linda, do you talk on the telephone with which ear? [Linda, with which ear do you talk (!) on the telephone?]

In example 10 there is main-clause question ordering (*How can she tell?*) in a subordinate clause—precisely the sort of ordering confusion one would expect the S procedure to be subject to. In example 11 S also seems to apply main-clause question ordering, as would have been correct in the yes/no question *Linda, do you talk on the telephone with your right ear?*. But this ordering cannot apply, because the *Wh* prepositional phrase has to appear in first position.

However, the possibility that the latter error has a conceptual cause should not be excluded. Cutler (1980a) showed that the intonation contour of this error is indeed that of a yes/no question. (Its prosody had been phonetically transcribed right after the error occurred.) It is likely that the speaker changed his mind after most of the question had been produced. A different sentence-initial PP (*with which ear*) was then erroneously accepted as a sentence-final complement of the running sentence. In that case, example 11 is a sentence blend, not an ordering error, according to Cutler.

The error in example 10 may also have an alternative interpretation. It is not different from what Garrett calls *shifts*—errors where closely adjacent words interchange. The words are typically not of the same syntactic category. Here is an example (from Garrett 1982):

(12) Did you stay up late VERY last night? [Did you stay up VERY late last night?]

Here the stressed word *very* is shifted over *late*. This cannot be a case where the categorial procedure (AP) applied a wrong but existing ordering scheme; there is no AP ordering that puts the adverb after the adjective. The error must have a different cause. Garrett proposed that shifts arise at a later level of processing, namely where the word forms are retrieved and assigned to the corresponding syntactical positions. In other words, they arise at the level of phonological encoding. It is hard, but perhaps not impossible, to find speech errors that can be exclusively adduced to a confusion of two legal syntactic ordering schemes.

There are other failures that should probably be interpreted as errors of grammatical encoding. Fay (1980a,b, 1982) discusses various such cases. One example, from Fay (1980b), is the following:

(13) Do you think it not works [*instead of* Do you think it doesn't work?]

In this case the VP of the embedded sentence failed to insert *do* in the presence of a negation. As a consequence, the present-tense parameter was realized as an inflection on the main verb (*works*).

Fay (1982), Stemberger (1985a), and others also discuss cases of sentence blends. Such cases seem to show that a speaker can, on occasion, encode two incompatible surface structures in parallel.

Speech errors clearly tell us something about grammatical encoding, but the evidence is in no way sufficient to give full-scale support to any one computational theory. And the situation is not different for the experimental evidence to be discussed in the following sections.

7.2 Units of Grammatical Encoding

In the previous section a distinction was made between call hierarchy and destination hierarchy. The former reflects the functional aspect of grammatical encoding: inspecting the message for functions and arguments to be encoded, retrieving the appropriate lemmas, and realizing the grammatical functions they require. The latter concerns the ordering aspect of grammatical encoding; it reflects which categorial procedures do the word and phrase ordering of the retrieved functional information.

From a computational point of view, one would expect a speaker's encoding process to be functionally driven in the first place. It is the function/argument structure of the message that the speaker is encoding in lemmas and grammatical functions. Ordering procedures, which are obligatory in the language, have no such expressive function. Ford and Holmes (1978) and Ford (1982) have obtained empirical support for the notion that the call hierarchy reflects a speaker's "planning units" more clearly than the destination hierarchy.

How can these two hierarchies be distinguished in a speaker's fluent speech? Ford and Holmes concentrated on the clause structure of the two hierarchies. Take, once more, the sentence of figure 7.1: *Attila believed the world to be flat*. In the call hierarchy (or "deep structure") there is a main and a subordinate sentence or clause. The subordinate clause starts after *believed*, so that the "deep" or "basic" clause partitioning of the sentence is like this (the beginning of a clause is indicated by a slash:

/Attila believed / the world to be flat (basic clause partitioning)

But there is only one S in the destination hierarchy, so the whole sentence consists of one "surface" clause (which is also called a "finite" clause, because it contains just one tensed or finite verb):

/Attila believed the world to be flat (finite surface clause partitioning)

It is generally possible to partition a speaker's fluent speech in either basic or finite clauses. The rule of thumb is that every finite clause contains one and only one tensed or finite verb (i.e., *believed* in the example sentence), whereas every basic clause contains one and only one main verb, whether tensed or not (i.e., both *believed* and *be*); there must always be a clause break between two main verbs. An example (from Ford and Holmes 1978) shows these finite and basic clause partitionings.

Finite clause partitioning:
/I began working a lot harder / when I finally decided to come to Uni

Basic clause partitioning:
/I began / working a lot harder / when I finally decided / to come to Uni

From the definition and from these examples it appears that the beginning of a tensed clause is always also the beginning of a deep clause, but not conversely.

If indeed the call hierarchy reflects a speaker's planning units, there should be evidence for basic clause partitioning in a speaker's delivery of a sentence. What sort of evidence? Ford and Holmes improved on a technique that Valian (1971) had developed for the same purpose. A speaking subject would occasionally hear a short tone, and his task was, apart from speaking, to press a button every time a tone appeared. The reaction time from tone onset to button pushing was the dependent measure. Speakers talked freely on various general topics, such as "family life," and tones appeared at unexpected moments but on the average once every 3.5 seconds. The expectation was that at moments when a speaker was engaged in planning the reaction times would be longer, because there would be less attention available for the impinging tone stimulus.

Ford and Holmes found increased reaction times for tones that occurred toward the end of basic clauses, even when the end of the basic clause was not also the end of a surface clause. A condition, however, was that more speech would have to follow the basic clause. Their interpretation of the data was that speakers start planning the next basic clause when they reach the final syllables of the current one. The data do suggest a basic clause rhythm in a speaker's fluent production. This testifies to the psychological reality of the call hierarchy.

Ford (1982) confirmed these findings by analyzing spontaneous hesitation pauses in speakers' fluent speech. She gave speakers essentially the same task as in the previous experiment, but without tones to react to. She then registered all hesitation pauses of 200 milliseconds and longer. She found that basic clauses were preceded by hesitation pauses as often as were finite clauses; pausing preceded about one-fifth of the clauses. Also, when there was pausing, the average pause length before the two kinds of clause was not statistically different. It amounted to about one second in both cases. Since the beginning of a finite clause is, by necessity, also the beginning of a basic clause, the simple rule seems to be that basic clauses "attract" hesitation pauses. It is irrelevant whether the beginning of the basic clause is also the beginning of a finite clause. In view of the earlier findings of Ford and Holmes, one might say that a basic clause starts being planned either in a preceding pause or concurrent with the uttering of the last syllables of the preceding clause (or both).

Ford also checked whether the *number* of basic clauses in a finite clause affected the duration of silence before that finite clause. A finite clause containing just one basic clause can be called a *simple* finite clause. A clause is *complex* when it contains more than one basic clause. If a speaker had been planning the full finite clause before initiating speech, simple finite clauses should have been preceded by shorter hesitation pauses than complex finite clauses. But there was no such difference. This finding supported the conjecture that speakers do not plan more than one basic clause ahead, even if the finite clause contains more than one basic clause.

Planning ahead does *not* mean that the whole clause is grammatically encoded before it is uttered. Experiments by Lindsley (1975, 1976), Levelt and Maassen (1981), and Kempen and Huijbers (1983) have shown that even the uttering of a simple subject-verb sentence can begin before the verb has been fully encoded. The only claim made by Ford and Holmes is that a speaker is occupied with the encoding of just one clause at a time. Only at the end of the uttering of a clause is the encoding of the next one taken up.

Ford concludes this way: "... the detailed planning for a sentence proceeds in recurring phases, each of which consists of the planning of a basic clause unit. During each phase a predicate is chosen and its lexical form retrieved, the logical arguments of the predicate decided upon, and the logical arguments sequenced in the way specified by the predicate's lexical form." We can agree, except for the last part. The sequencing or ordering of phrases is determined not only by the structure of lemmas and the call hierarchy but also by the destination hierarchy. The major categorical procedures impose order on the returned constituents. The same lemma can find its arguments arranged in different orders, depending on such factors as whether it occurs in a main clause or a subordinate clause.

Taken together, these findings are supportive of the notion that the rhythm of grammatical encoding follows the semantic joints of a message —its function/argument structure—rather than syntactic joints. It is the partitioning of the message to be expressed that a speaker is attending to, and this (co-)determines the rhythm of grammatical encoding. The syntactic-ordering decisions are largely or fully automatic; they can run in parallel. Characteristic speech errors may occur when operations of syntactic ordering fail, but the complexity of these syntactic operations will not be reflected in measures of mental load, such as reaction times and hesitation pauses. This is in line with earlier studies by Goldman-Eisler (1968), Taylor (1969), and Rochester and Gill (1973), who found no effects of syntactic complexity on hesitation pauses, speech latencies, or disruptions, respectively.

Gee and Grosjean (1983) performed extensive analyses of hestitation pauses, relating them not only to clause boundaries but also to any other phrase boundaries. They also included *phonological* units in their analysis, in particular so-called intonational and phonological phrases. Prosodic units and their generation will be discussed in section 8.2 and chapter 10, but it is important here to notice that these phonological units added substantially to the account of the observed pausing patterns. Van Wijk (1987), in a reanalysis of the Gee-Grosjean results, argued that the observed pausing patterns are *exclusively* determined by the phonological encoding and the articulatory preparation of the speaker. In other words, grammatical encoding plays no role worth talking about.

Does this contradict the above conclusions on deep clauses as planning units? Not necessarily. First and foremost, the data of Gee and Grosjean were obtained on a *reading* task, whereas the data of Ford and Holmes involved spontaneous speech. The course of grammatical encoding is probably quite different in these two kinds of task. In reading, the speaker can rely heavily on the printed materials. Lexical retrieval and the building of syntactic constituents can be based largely on parsing of the visual input. Reading aloud is primarily a perceptual, phonological, and articulatory task. It is not at all surprising, then, to find no effects of message planning and its interfacing with grammatical encoding (i.e., with the call hierarchy). Second, the original Ford-Holmes data were reaction times, not pause patterns. There is no compelling reason for a phonological or an articulatory interpretation of such reaction-time data. Third, the issue is not one of either/or. It would, in fact, be quite surprising if processes of phonological encoding and articulatory preparation were *not* responsible for generating pause patterns; they are the proximal causes of pausing. What is at issue is whether, in addition, there are traces of more distal encoding operations. Mapping function/argument structures in the message onto clause-like syntactic units probably leaves such a trace in the pause pattern. Subsequent findings by Holmes (1984 and forthcoming) are consonant with this conclusion.

7.3 The Encoding of Topic and Other Nuclear Entities

7.3.1 Accessibility and The Encoding of Topic

As was discussed in section 3.4, a speaker takes a certain perspective on a conceptual structure to be expressed. There is, first, the choice of topic. When the speaker's purpose is to expand the addressee's knowledge about something, the message will highlight this topic concept, to distinguish it from the comment that is made about it. In its turn, the Formulator will encode the topic in a syntactically prominent position.

What does "syntactically prominent" mean? It can mean that the topic is encoded as grammatical subject. Remember that there is a preference hierarchy for grammatical functions, with the subject function at the top of the list (see subsection 6.2.2). It can, alternatively, mean that the topic will be encoded early in the sentence, whether or not in the role of subject. This makes good functional sense: When the speaker's intention is to expand the interlocutor's information about something, the interlocutor may want to first find or create the address to which the comment information can be attached. This is easier when the topic information appears early in the sentence, before the verb, than when it appears after the verb.

It is very generally the case that these two carriers of syntactic prominence coincide. Word-order statistics of the world's languages show that 84 percent of them place the subject in front of verb and objects (Hawkins 1983). Hence, when a subject function is chosen to encode topical information, it will usually precede the predicate containing comment information. But, as we will see shortly, it may be possible to distinguish empirically between topic "fronting" and assigning subjecthood to topics.

How does the Formulator manage to place the topic in a syntactically prominent position? I will argue that this is an automatic consequence of the Formulator's control structure. It will, in many cases, be enough for the topical concept to be first in triggering its lemma(s) for it to be encoded early in the sentence and/or in subject position. This is exactly what happened to the concept CHILD in the procedural example of section 7.1, where the sentence *the child gave the mother the cat* was generated. The concept CHILD was an early-available message fragment. Maybe the speaker had introduced it in the previous sentence, so that the concept was highly available. At any rate, it caused *child* to become the first-retrieved lemma. The further operations, which generated *the child* as a subject noun phrase in sentence-initial position, were automatic consequences of this early availability of the lemma. The early triggering of topical lemmas could, in many cases, be due to an early delivery of the topical message fragment. Wundt's principle will then automatically give it primacy in grammatical encoding as well.

Bock and Warren (1985) called this factor "conceptual accessibility." They showed that a highly available concept tends to be encoded in a prominent grammatical function. If this is the factor at work, it should not really matter whether it is topichood that makes a concept highly available. Alternatively, or in addition, there may be a preliminary encoding operation by the categorial procedure S, which checks for the fragment in the message that is explicitly marked as topic. Theories are lacking here, and we now turn to some empirical evidence showing the preference of topical or otherwise highly accessible information for sentence-initial position and for subjecthood.

The experimental work in this context has been mostly concerned with the elicitation of the passive voice. Subjects asked to describe simple events where there is an agent doing something to a patient or recipient (for instance, a dog attacking a cat) prefer the active voice (i.e., a sentence like *the dog is attacking the cat*). Agents are preferably expressed in subject position. The simplest explanation for this fact is that the subject who perceives such an event normally encodes it from the perspective of the agent (i.e., it is the dog who is doing something). In order to transmit this information to the listener, one should make the dog the topic so that the addressee can know that the predication is about the dog. The topic (i.e., the agent) is then grammatically encoded as subject, in any of the ways discussed above. If this is what happens, experimental manipulations that make the patient (the cat) topical should increase the probability that the speaker generates passive voice. Why? If the cat is topic, it will tend to be expressed sentence-initially in subject position, and this will favor the use of the passive verb lemma (subsection 6.2.2), since it is the one that maps the thematic patient onto the subject function: *the cat is attacked by the dog*.

Various types of experimental manipulation have been used to topicalize the patient, or at least to make it more accessible. A first method is to give a perceptual cue. One shows a picture of the patient either before or after presentation of the scene, but before the subject gives his description. And the control condition, of course, is to show a picture of the agent. Prentice (1967) and Turner and Rommetveit (1968) used this method and obtained the predicted results. If the agent was cued (e.g., by a picture of the dog), almost no passive responses were given; however, if the cue depicted the patient, recipient, or theme, there were a significant number of passives in the descriptions (such as *the cat is attacked by the dog*).

Other researchers gave verbal cues. Tannenbaum and Williams (1968) first presented a paragraph in which either the patient, or the agent, or neither figured centrally, and this was then followed by the scene to be described. The subject's task was either to describe the scene with an active sentence or to describe it with a passive sentence, and the time needed to complete the sentence was measured. This is a rather unnatural task, but there was a clear congruency effect in the data. Both actives and passives were produced relatively quickly when the subject of the sentence agreed with the topic induced by the preceding paragraph. Flores d'Arcais (1975) simply mentioned either the agent (e.g., the dog) or the patient (e.g., the cat) before presenting the picture of the event (e.g., a dog attacking a cat). The subjects (native speakers of Italian) were free to describe the scene with either active or passive sentences, but the verb *attack* (in Italian) had to be used. When the cue word mentioned the agent, 77 percent of the responses were active sentences; when the cue referred to the patient, 67 percent of the responses were passives.

A very natural way to induce a topic is by asking a question. The subject is asked a question about the scene—for instance, "What did the dog do?" or "What happened to the cat?". Here the interlocutor indicates which protagonist information is needed, i.e., which protagonist should be topic. This manipulation was used by Carroll (1958), who performed the first experimental study of the elicitation of passives; it was also used in a study reported in Bates and Devescovi (forthcoming). In the Bates-Devescovi study, subjects were shown short filmstrips, for instance about a hippo hitting a ladder. They were then asked about entities in the film (e.g., "Tell me about the hippo" or "Tell me about the ladder"). When the request focused on the agent, 100 percent of the responses were active sentences; when the patient was focused, 70 percent of the answers were passive sentences (e.g., *the ladder was hit over by the hippo*). Bates and Devescovi also measured response latencies, and found a clear congruency effect. Active sentences have relatively short latencies; however, when the patient is probed and the subject still uses an active sentence, the latency is relatively long, and longer than the latency for a passive response.

Question asking was also used in a study by Bock (1977). She used a recall task. Subjects were first presented with a list of sentences of various syntactic structures. They were then given a question as a recall cue for one of the sentences on the list. The questions were constructed in such a way that they would topicalize a particular entity in the sentence to be recalled. If the sentence *a psychologist cured a neurotic poodle* had been on a subject's list, the following question could be asked: *The interior decorator was afraid she would have to get rid of her neurotic pet poodle because it was ruining the furniture, but she was able to keep it after all. What happened?* Subjects tended to reproduce the sentence in the passive voice: *a* (or *the*) *neurotic poodle was cured by a psychologist*. Clearly, the context sentence here is about the poodle, not about the psychologist, and in the answer the poodle is encoded sentence-initially in subject position.

Another way to elicit passives in recall tasks is to give a recall cue relating to the patient of the memorized sentence. Variations on this procedure can be found in Prentice 1966, Turner and Rommetveit 1968, Perfetti and Goldman 1975, and Bock and Irwin 1980. Bock and Irwin, for instance, would read a list of sentences to a subject, and next prompt the sentences by way of key words. If the sentence had been *The falling tree crushed the lumberjack*, the prompt word could be either *tree* or *lumberjack*. The results of the experiment were quite clear: The reproduced sentence tended to have the constituent containing the key word in a relatively early position. When *tree* was key word, subjects would reproduce the sentence in its original form. But when the key word was *lumberjack*, many reproductions took the form *The lumberjack was crushed by the falling tree*. Similar findings were obtained for other sentence pairs that differed in word order.

The linkage of sentence-initial position and subjecthood, we saw, is not absolute. A topic or a highly accessible entity can be encoded early in the sentence without becoming a subject. This is harder in English than in languages that have freer word order. To describe a scene where some man throws some ball, a German speaker can topicalize the ball by encoding the scene as *Den Ball wirft der Mann*, where the ball is fronted (and accented) without being subject; it has accusative case. In English it is not so easy to disentangle fronting effects from the assignment of subjecthood. But Bock and Warren (1985) developed an experimental paradigm that did just that: a sentence-recall task in which the main verb of the sentence was used as the recall cue. A subject would listen to a list of sentences. In this list there were sentences taken from three binary sets. Examples are given in table 7.1.

Table 7.1
Kinds of sentences used in Bock and Warren's (1985) recall task.

Type 1	**Actives versus passives**
Active:	The doctor administered the shock
Passive:	The shock was administered by the doctor
Type 2	**Prepositional versus double object**
Prepositional:	The old hermit left the property to the university
Double object:	The old hermit left the university the property
Type 3	**Natural versus unnatural phrasal conjuncts**
Natural:	The lost hiker fought time and winter
Unnatural:	The lost hiker fought winter and time

The two alternatives within each type always involved a different ordering of two nouns: *doctor* and *shock* in the type 1 examples, *property* and *university* in the type 2 examples, and *time* and *winter* in the type 3 examples. For types 1 and 2 these ordering alternatives expressed a different assignment of grammatical functions. In the type 1 examples, either *the doctor* or *the shock* would be subject of the sentence. In the type 2 examples, *the university* would be either prepositional object or indirect object. But no such difference in grammatical function is apparent in the type 3 alternatives—the two nouns have the same function in either order; one can only say that an order in which the shorter word comes first sounds somewhat more natural (Cooper and Ross 1975). The first question Bock and Warren asked themselves was this: When the subject recalls the sentence, will there be a tendency for the conceptually more accessible entity to be projected on a more prominent grammatical function (i.e., subject for the type 1 sentences and indirect object for the type 2 sentences)? The second question was: Will the more accessible entity tend to be expressed in a more fronted position? Only in this latter case will there be an effect for type 3 sentences. If, however, the ordering of elements is mediated only by grammatical function, there will be an effect for sentences of types 1 and 2 only, not for those of type 3.

How was the accessibility of the entities manipulated? Through imageability (Paivio, Yuille, and Madigan 1968). One of the two critical nouns in the sentence was high on imageability (according to the tables of Paivio et al.), and the other one was low on imageability. In half the sentences presented to a subject, the highly imageable noun preceded the less imageable one; in the other sentences, the order was reversed.

In the recall task, the subject was given a sentence's main verb (e.g., *administered*) as a cue for reproducing the sentence. Regrettably, the subjects were asked to write down the sentence instead of to speak it; one should be careful in generalizing the results to normal speaking conditions, where prosody could play a significant role. But the results were clear cut. There were substantial effects of imageability for the type 1 and type 2 sentences, but none for the type 3 sentences. Bock and Warren concluded that conceptual accessibility does not affect ordering directly. Rather, a more accessible entity will "claim" a more prominent syntactic function. This, in turn, may result in a more frontal syntactic position.

Clearly, Bock and Warren's method did not involve explicit topicalization of entities. Neither were any questions asked about the critical referents (such as *What did the doctor do?*), nor did they figure as cue words (as in the Bock-Irwin study). Hence, it is still possible that topicalization can be done by "mere" fronting, without assignment of subjecthood to the topical entity. In English this would require a clefting construction, as in *It is the ball that the man throws* or *As for the ball, the man throws it*. But the latter construction, especially, has the additional effect of suggesting that there is a *change* of topic (see Reinhart 1982 and the references given there). Sridhar (1988), in a cross-linguistic study, found that in many languages, such as Hungarian, subjects topicalized objects by fronting them as in the German example above.

But if this is possible, why would speakers front a topic constituent without at the same time encoding it as the grammatical subject of the sentence? One reason may be that there is a strong competitor for the subject function. Actors and agents are preferably encoded as subjects of sentences, and more often than not they will be topics as well. But if the situation (e.g., a question) requires another argument to be the topic, the conflict can be resolved by adapting the word order and keeping the canonical assignment of grammatical function. This is what happened in the German example above. The agent (*der Mann*) is still in subject role, but the ball is placed in frontal position. Reinhart (1982) put it this way: "subjects are the unmarked topics, which means that it is easier to use a sentence when we intend its subject to be a topic. But they are not obligatory topics." The notion of competition is central to the so-called competition model of MacWhinney and Bates, which will be taken up in the next subsection.

7.3.2 Encoding Nuclear and Non-Nuclear Entities: Saliency and Competition

The notion of congruency was used in the previous paragraph to express a preferred relation between message and syntactic form. A message with a topicalized recipient has a preferred (though not a necessary) relation to a passive syntactic form. Whether that grammatical encoding will obtain depends on a variety of other pressures in the encoding process—for instance, pressure to assign the subject function to a salient agent or to an otherwise prominent or highly accessible conceptual entity. Subsection 6.2.2 discussed the notion of a universal hierarchy of grammatical functions, going from subject via direct and indirect objects to obliques. Arguments "want" to be encoded through the highest-ranking function possible. And there is at the same time a pecking order of thematic roles: from agent, via theme, to source, goal, and other roles. The agent is usually encoded in the most prominent grammatical function (i.e., as the subject); source and goal generally end up in oblique prepositional phrases. A congruent grammatical encoding is one in which the arguments and modifiers in the message are distributed over grammatical functions and syntactic positions in a way that reflects their relative importance.

But what determines the saliency or importance of an entity in the message? Several factors have already been reviewed. There were discourse factors, such as whether the entity had been topicalized by an interlocutor's question, or whether it had been recently mentioned. We also saw that imageability played a role. However, a host of other factors may be involved as well. They all relate to what can be called the "human interest" of an entity—its relevance in the eye of the speaker. This is, admittedly, rather vague, but the literature contains several suggestions as to what might contribute to human interest or to the foregrounding of certain concepts. Fillmore (1977) talked about a *saliency hierarchy*. Among the factors contributing to saliency is *humanness*. Human concepts have a stronger tendency to be encoded in the major grammatical functions subject and object than nonhuman concepts. Another factor Fillmore mentioned is *change of state*. It is usually of more interest that a state changes than that it does not change. A speaker will foreground a changing object in the message, and hence increase the odds that it will be encoded as subject or object. Compare examples 14 and 15. In the former, the table is not moving and is encoded in an oblique grammatical function; in the latter it is moving and is encoded as direct object:

(14) I pushed against the table

(15) I pushed the table

Still another factor contributing to saliency, according to Fillmore, is *definiteness* (or *accessibility*, in the terms of subsection 4.5.1). A definite referent will attain a prominent grammatical function more easily than an indefinite one. In describing a scene where there are two boys and one girl, it is probably more natural to say *The girl is kissed by a boy* than *A boy is kissing the girl*, in spite of the fact that the boy is depicted as the agent.

Chafe (1977) and Clark and Clark (1977) speak of *unexpectedness* as contributing to saliency or information value. Osgood and Bock (1977) show that *vividness* is a contributing factor. MacWhinney (1977) mentions *agency*, as well as other factors. And *perceptual factors* have often been mentioned in the literature. Figures are usually more salient than grounds, closer objects more salient than faraway ones, bigger objects more salient than smaller ones, and so on.

It should not be the aim here to review the varieties of human interest. All that is at issue is the claim that foregrounded, nuclear, emphasized entities in the message typically find their grammatical encoding in higher grammatical functions or earlier in the sentence than backgrounded or nonnuclear entities. What is the empirical evidence?

Surprisingly, there is only fairly limited evidence from analyses of (elicited) spontaneous speech. Most studies have resorted to asking subjects which sentence is a more or a less appropriate one given a particular state of affairs, asking them to put printed words or constituents in some natural order, and so on. Osgood's seminal study (1971; see also Osgood and Bock 1977), and a substantial expansion of it by Sridhar (1988), are closer to spontaneous speech. In these experiments subjects were asked to describe, one by one, a sequence of short scenes or events. In Osgood's study these were acted out (by Osgood); in Sridhar's experiment they were filmed scenes. A disadvantage of both studies is that the subjects wrote rather than spoke their descriptions of the scenes. In this way the subject could not use prosodic ways of expression, and may have compensated by making another choice of syntactic forms. Still, these studies have provided interesting information about the grammatical encoding of saliency.

Osgood played a sequence of 32 events on a table in front of an audience consisting of 26 students. The events involved a set of simple familiar objects, such as various balls of different size and color, a plate, a tube, and some chips. The subjects were instructed for each demonstration to open their eyes on the signal "open" and to close them on the signal "close". This was followed by the request to describe the event they had observed in a single sentence that "a six-year-old boy 'just outside the door' would understand." They were also instructed to describe the actor (Osgood), when necessary, as "the man."

A typical event observed by the subjects was one in which an orange rolling ball is hitting an upright tube. What would be a congruent description? The ball is agent and is changing location; the tube is recipient and static. This would favor encoding the ball in subject position. On the other hand, the tube had also figured in the previous demonstration, whereas the ball had not been around for some time. This might favor making the "given" tube the topic of the sentence. Twenty-one subjects produced sentences where the ball was subject, and five gave passive sentences with the tube in subject position.

The effect of humanness or animacy was apparent all over. When Osgood held a not-previously-observed big blue ball, 20 subjects made sentences where *the man* was subject (like *The man is holding a big blue ball*). Only five found the saliency of the new big ball impressive enough to make a passive sentence (such as *A big blue ball is held by the man*). When "the man" was seen to be involved in the scene, either in a static or in an agentive role, he usually turned up in subject position.

Figure/ground organization was also very effective. One would expect figures to be encoded in more prominent grammatical functions than grounds. When the scene was just a ball on the table, nobody put the table in subject position (something like *The table is supporting the ball*). It was encoded obliquely (*on the table*) or not at all. In describing a more complicated static scene, where a ball was on an upright tube and the tube stood on a plate, subjects normally arranged the noun phrases such that the ball would be first, the tube second, and the plate third. The ball was usually encoded as subject, and the plate obliquely in a prepositional phrase.

Osgood also played saliency factors off against one another—for instance, change of state and vividness. When he presented a scene where a very big orange ball (vivid) was hit by a small black ball (change of state and agency), only three subjects created a passive sentence with the big ball in subject position. It has been a general finding (see also Sridhar 1988 and Bates and Devescovi, forthcoming) that passives are quite hard to elicit by purely perceptual means. They are primarily triggered by discourse constraints.

Sridhar (1988) filmed Osgood-type scenes and events and presented them to subjects from ten different language communities: Cantonese, Hebrew, Finnish, Hungarian, Serbo-Croatian, English, Spanish, Japanese, Kannada, and Turkish. This cross-linguistic approach made it possible to test the universality of the congruency effects we are discussing—in particular, the claim that more important, salient, or informative entities will be encoded in major grammatical functions, and relatively early in the sentence.

Sridhar obtained overwhelming support for this claim. When, for instance, the scene was a man rolling a ball on the table, the man and the ball were almost invariably encoded as subject and object, respectively, while the table, if mentioned at all, would never appear in a major grammatical function. When there was a figure and a ground, the figure was encoded earlier in the sentence than the ground in 70 percent of the responses, and this tendency was apparent in all languages. Actions were encoded more often and earlier in the sentence than changes of state, and changes of state more often and earlier than constant states. Descriptions of static scenes proceeded preferably from top to bottom, as in Osgood's ball-on-tube-on-plate example (which was included in Sridhar's film). Closer-to-ego objects were related to further-away objects, and appeared earlier in the sentence (similar evidence was presented in subsection 4.5.3 above).

Sridhar also included scenes where (quasi-)humanness could be tested without confounding it with agency. In one scene, a black ball hit a doll which was sitting on the table. The doll appeared before the ball in 67 percent of the descriptions. When, however, a scene was presented in which the black ball hit a yellow ball, only 43 percent of the descriptions had the yellow ball before the black ball. This pattern of promoting a human entity, even if it is only a thematic patient, held for 9 of the 10 languages (Kannada was the exception).

These and many more results of Sridhar's study demonstrate that speakers of vastly different languages show very similar congruency mechanics. Materials that are presumably foregrounded or "in perspective" in the message are grammatically encoded in major grammatical functions; and the more important they are (by a variety of importance criteria), the higher they climb in the grammatical-function hierarchy and the earlier they appear in the sentence.

In one study in this tradition, spoken language was elicited. Flores d'Arcais (1987b) presented moving geometrical figures of various shapes to subjects and asked them to describe each such event. When a big figure and a small one were involved in an event, the big figure was mentioned first. When one figure was leading the other in the movement, the leading figure was mentioned first. Such factors determine which entity becomes the topic of the event.

As was observed above, the different forces from the message can often be in conflict. The topic entity may not be the agent; a vivid and brand-new object may not be more than the background for an action; and so on. Such competitions must be resolved in fluent speech. In addition, different languages provide different means for resolving these conflicting tendencies. In free-word-order languages the assignment of grammatical function is relatively independent of constituent order, with nonconfigurational languages such as Malayalam (see subsection 5.1.1) as extreme cases. Speakers of such languages may resolve conflicts by mapping one message function (e.g. topicality) onto word order (sentence-initial) and another (for instance, agency) onto grammatical function (subject). But this is far less easy for speakers of languages with rather fixed word order, such as English.

Bates, MacWhinney, and co-workers (see especially Bates and MacWhinney 1982, 1987; MacWhinney, Bates, and Kiegl 1984; Bates and Devescovi, forthcoming) have proposed a "competition model" in which a variety of message-level aspects, including thematic structure, topicality, and foregrounding, are probabilistically mapped onto a variety of forms at the syntactic level. The probabilistic parameters of this mapping are assumed to be different for different languages. An effort is made to find out how the parameters—i.e., the preferred use of particular syntactic devices for particular aspects of the message—can be understood from the structure of the language. Though most of this work is still based on sentence-interpretation data, data on spoken production have been acquired and analyzed (Bates and Devescovi, forthcoming) by a method roughly like Sridhar's film-presentation procedure.

The competition model is claimed to describe a *direct*, though probabilistic, mapping between the functional message level and the formal grammatical level. In other words, the congruency mechanics consist of direct grammatical responses to message "stimuli." This is acceptable if "direct" doesn't mean more than that grammatical encoding is an automatic modular process, not involving intentional goal-directed choices on the part of the speaker. There must, however, be mechanisms that realize the mapping. The "direct" probabilistic mapping of the competition model thus means that the mechanism is inherently probabilistic. In other words, an individual language user's Grammatical Encoder is supposed to be a probabilistic processor.

This is surely a respectable option, in line with an equally respectable tradition of stochastic modeling in psychology. However, the probabilistic data on which the model is currently based do not require such a solution. These data are averages over subjects, not within subjects. It could very well be the case that each individual language user has a consistent *deterministic* way of encoding messages into surface structures. The probabilistic findings are then entirely due to processing differences *between* subjects. In short, the data underlying the competition model do not relieve us from the task of developing a mechanical (computational) model of grammatical encoding or from the task of checking whether a deterministic model (with parameters differing between subjects) might suffice.

7.4 Cohesive Encoding

Another important aspect of a speaker's grammatical encoding is the realization of what is commonly called *cohesion* (Halliday and Hassan 1976). In cohesive discourse the speaker makes, where necessary, the form of the current utterance dependent on what was previously said by himself or the interlocutor. This makes it possible for the addressee to interpret elements in the current utterance in relation to previous discourse (or, occasionally, to the discourse that will immediately follow). But the cohesive forces are not limited to the ways in which entities are referred to; they also permeate the speaker's syntax. In the following, some attention will be given to empirical studies of both kinds of cohesion.

7.4.1 Cohesive Reference

What grammatical means can a speaker use to establish and maintain reference? In chapter 4 we began paying attention to the ways in which a speaker can signal to the addressee what the accessibility of a referent is, given the state of the discourse model. It was suggested that the speaker marks each referent in the message in a triple way, depending on the referent's relation to the discourse model. The marking will be plus or minus "accessible", plus or minus "in discourse model", and plus or minus "in focus" (see subsection 4.5.1). Each of these markers, when correctly encoded in the surface structure, contributes in its own way to directing the attention of the listener, making it possible for her to find or create the appropriate address for the referent in the discourse model.

How are these accessibility markers grammatically encoded? Many languages encode the "accessible" marker in the noun phrase's morphology or determiner, mapping " + accessible" referents onto definite noun phrases and " − accessible" ones onto indefinite noun phrases. The index "in discourse model" is often mapped onto prosodic focus. A referent which the speaker supposes to be in the interlocutor's discourse model is then deaccented. The "in focus" feature is often realized by some sort of lexical reduction. In-focus elements tend, in particular, to be pronominalized; there can even be complete elision. But languages also have other means to encode the referent's accessibility status, such as by word order or special morphology.

Experimental and observational support for the appropriate encoding of these anaphoric features in English were reviewed in chapter 4. Here we will limit our attention to some cross-linguistic results which support the notion that these features are universally encoded by speakers. Also, there are certain commonalities between speakers of different languages in the way this grammatical encoding is realized, but there are interesting differences as well.

Osgood's (1971) study contained rich data with respect to the circumstances of definite and pronominal reference. Sridhar's (1988) work showed that Osgood's main findings also hold for languages other than English. Osgood had observed that an object introduced for the first time was described with an indefinite article in 85 percent of the cases, but when it appeared a second time the percentage dropped to 45. Sridhar also found his subjects referring to new entities by way of indefinite noun phrases, but in some languages such an indefinite NP was not very acceptable as the subject of the sentence. This was especially the case for Hungarian and for SOV languages (i.e., languages that normally have the verb after the subject and the object). In particular, the Japanese subjects evaded ordering the indefinite term before the definite one. Notice that also in English an indefinite NP is not an exciting start of a sentence (*A ball is on the table*), in agreement with Fillmore's (1977) observations. In so-called presentative contexts, such as this one, many languages favor pre-posing the verb (*There is a small ball on the table*).

With respect to reduction as a means to express that a referent is in fcous, Osgood had found that pronominalization was especially frequent when the same object participated in two events during the same demonstration. For instance, when a black ball hit a blue one, and the blue one in turn hit an orange one, the second mention of the blue ball (now in the interlocutor's focus) was mostly by means of a pronoun (as in *the black ball hits the blue ball. It hits the orange ball*). Osgood had also found a sharp decrease in the number of adjectives when the same object was described a second or a third time (*a big blue ball* → *the blue ball* → *the ball*). Sridhar confirmed this tendency for all of his ten languages, though the tendency was slight in the responses of Japanese subjects.

MacWhinney and Bates (1978) reported a systematic cross-linguistic study of spontaneously produced speech. Children of different ages as well as adults were given pictorial presentations of scenes to be described. There were three language groups: native speakers of Italian, English, and Hungarian. With respect to the feature "accessible", the finding was that English and Italian speakers made use of the indefinite article to express inaccessibility. But this was much less the case for the Hungarian speakers —presumably because in Hungarian a new referent is introduced without any article (this also explains Sridhar's finding for Hungarian, mentioned above). The English speakers were less fond of using the definite article than the Italians and the Hungarians, probably because the indefinite article has an additional function in the latter two languages (namely as a numeral). Subjects apparently used the definite article wherever they could, in order to evade the potential confusion of functions.

In the MacWhinney-Bates study, the borderline between full ellipsis and pronominalization turned out to be different for English and the other two languages. Where English speakers used a pronominal form in subject position, Italians and Hungarians simply elided the subject. This is mostly impossible in English. The person and number of the elided subject are clearly marked in the verb morphology of Italian and Hungarian, and hence there is no loss of information in those languages.

What about pitch accent in the MacWhinney-Bates study? Newly introduced elements were given pitch accent, as in the studies of Nooteboom and Terken (1982) and Pechmann (1984) discussed in chapter 4. But there was a substantial difference between languages. English-speaking subjects used much more prosodic stress to introduce new or contrasting entities than did Italian speakers. And the Italians stressed more than the Hungarians, who hardly used any pitch accent for the introduction of new referents. It seems that the latter two languages offer more possibilities to use word order for marking the introduction of a new element. And indeed, there was much more fronting of "new" constituents in Hungarian and Italian than in English. This gives support to the idea that the feature of being new in the discourse model is not necessarily expressed by prosodic means. Different languages can provide different grammatical encodings for this feature.

7.4.2 Cohesive Syntax

Speakers are also cohesive in their wording and syntax. Some of the evidence was reviewed in subsection 4.2.5, in particular the observational work by Schenkein (1980) and the experimental study by Levelt and Kelter (1982). The latter study showed the speaker's tendency to answer a question like *(At) what time do you close?* with an answer that was congruent in the use or nonuse of the preposition (at). These repetition effects are probably not due to the content of the answer.

Strong experimental evidence for the persistence of syntactic form was collected by Bock (1986b). We will call her experimental method the *priming/recognition procedure*. It was also profitably used for other purposes, to which we will return in the next subsection. The experimental manipulation is one of priming, but the subject is made to believe that it is a recognition test. Bock's study involved syntactic priming. The speaker was to describe a set of pictures, but each picture was preceded by a (prime) sentence, spoken by the experimenter and repeated by the subject. The question, then, was whether the syntax of the prime sentence would affect the syntax of the picture description. The camouflage of the experiment consisted in first giving the subject a "study list" also consisting of pictures and spoken sentences. Only after this was the experimental list presented, and the subject was made to believe that this was a recognition task. The instruction was to say "Yes" or "No" for each sentence or picture, depending on whether that item had appeared in the study list. But before giving that answer, the subject had to repeat the sentence or describe the picture.

The prime sentences used in Bock's study were of the same kind as types 1 and 2 in table 7.1. That is, they were actives or passives, and prepositional or double-object sentences. The pictures (examples of which are given in figure 7.2) invited either type 1 descriptions (*Lightning is striking the church*, or *The church is being struck by lightning*) or type 2 descriptions (*The man is reading a story to the boy*, or *The man is reading the boy a story*).

Bock's results were unequivocal. The picture descriptions tended to be syntactically congruent with the prime sentences. If the prime was passive,

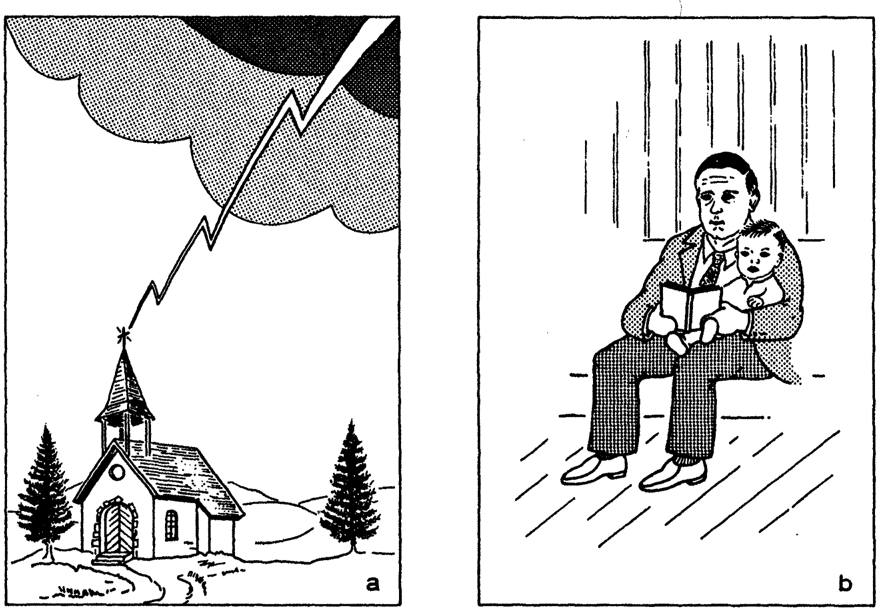


Figure 7.2
Examples of pictures used by Bock (1986b) to elicit scene descriptions.

the description tended to be passive; if the prime was a double-object sentence, the picture description tended to be a double-object sentence, and so on.

How does such syntactic coherence arise? There is nothing in the algorithmic theory of section 7.1 that would predict this result. On that theory, syntactic variation can only result from varying the order in which message fragments are made available. But that was not at issue in Bock's experiment. As in the Levelt-Kelter experiments, it turned out that variations in grammatical encoding could be induced by other than message-level means. In Bock's experiments this was probably not a (sole) consequence of lexical priming. For instance, when the prime was a full passive, the description could be a truncated one (like *the church was struck*)—the word *by* was not repeated, but passive syntax was.

An interpretation that stays close to the Kempen-Hoenkamp algorithm is that the syntactic specialists—in particular, the categorial procedures S, VP, NP, AP, and PP—can be biased by listening to and repeating a sentence of a particular syntactic form. This pleads for the independent existence of these syntactic specialists. They can be biased in such a way that grammatical encoding operations *overrule* message-level factors.

7.5 Feedback in Grammatical Encoding

Conceptual accessibility, we saw, has an effect on grammatical encoding. A message entity that is early available or otherwise particularly salient will normally succeed in claiming a relatively prominent grammatical function. The most likely explanation for this finding is that the corresponding lemma is retrieved at an early stage. The NP procedure called by that lemma (if it is a noun or a pronoun) delivers its output according to some preference order. The default claim is subject-of-S. If that slot is already occupied, direct-object-of-S is the next preference, and so on down the grammatical hierarchy. In this way, the order of conceptual activation can affect the order of grammatical encoding. The order of grammatical encoding, in turn, affects the order of phonological encoding. The appropriate word form can be retrieved as soon as its diacritic parameters (for number, case, etc.) have become available, and this depends, in part, on the grammatical function assigned. So we see Wundt's principle at work at two levels: from message planning to grammatical encoding, and from grammatical to phonological encoding.

The present section discusses whether there is any *converse* effect. Is there evidence for feedback in the system such that (a) the accessibility of lemmas can affect conceptual ordering decisions or (b) the accessibility of word forms can affect the order of grammatical encoding? It should be remembered that the research strategy proposed in chapter 2 was to take nonfeedback as our starting hypothesis. The assumption of "informational encapsulation" was that processors would be sensitive only to their own characteristic kind of input, not to feedback from other components. In this case, conceptual decisions should be ignorant of the results of grammatical encoding, and grammatical encoding should be insensitive to what happens at the level of phonological encoding. But this is an empirical issue. Let us now review some of the supporting and contradicting evidence.

The first issue is whether conceptual ordering decisions can be affected by lower-level formulation processes. This question was put to test by Levelt and Maassen (1981), who tested, in particular, whether the relative accessibility of lexical items affected the order of mention in conjunctive constructions. The starting assumption was that the order of mention is a conceptual message-level decision. Subjects in these experiments were presented with pictures containing three geometrical forms and a fixation point. An example is given in figure 7.3. Immediately upon presentation of the picture, two of the three forms (for instance, the circle and the square) were slightly displaced upward or downward. The subject's task was to say as soon as possible what had happened. Depending on the directions of motion, typical responses were *the circle and the square went up* or *the circle went up and the square went down*. The former sentence type, labeled *noun-phrase conjunction*, was given only when the two forms moved in the

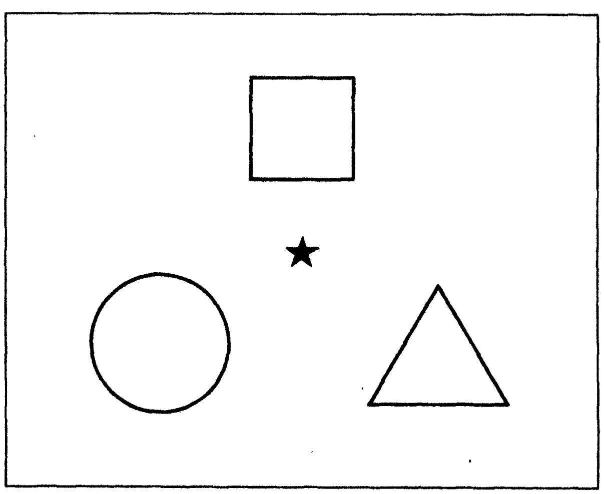


Figure 7.3
Example of the stimuli used by Levelt and Maassen (1981) to elicit two-object event descriptions.

same direction: both up or both down. The latter type, called *sentence conjunction*, was always used when the directions of movement were different. Sometimes subjects used sentence conjunction when the movements were in the same direction (*the circle went up and the square went up*), but this was less frequent and surely less congruent since it doesn't do justice to the "common fate" of the two motions. (In an additional experiment, the sentence forms were prescribed rather than left to the subjects' convenience.)

The geometrical figures had been carefully selected. Half of them had to be easy to name and the other half hard to name, and this difference had to be due to lexical accessibility rather than to better or worse recognizability or detectability. The figures were selected in preliminary experiments where subjects simply had to name individual geometrical patterns, and where naming latencies were measured. The figures indeed differed markedly in naming latency. In addition, the detectability latencies were measured for all figures. Eventually, sets of easy- and hard-to-name figures were selected that did not differ in detectability. A circle, for instance, was an easy-to-name figure, with an average naming latency of 691 milliseconds. The naming latency for *square*, however, was relatively long: 772 milliseconds on average. Hence, the square was in the hard-to-name set of figures. The experiments involved three easy-to-name and three hard-to-name figures. Their average difference in naming latency was 83 milliseconds.

Did lexical accessibility affect word order? When the event involved a circle and a square, for instance, did the speakers say *the circle and the square went up* (where the easily accessible word *circle* is in initial position), or *the square and the circle went up* (with the harder word first)? Notice that the order "easy → hard" was quite cooperative, given the task. The subjects had been asked to give their descriptions as quickly as possible. By uttering the easy word first, the speaker would gain in descriptive speed.

But nothing happened. The easy word was first in 52.5 percent of the descriptions—not significantly different from randomness. (In the additional experiment with prescribed syntactic response frames, the percentage was 50.5.) What *was* found was that when the hard word was put in initial position, the latency to initiate the sentence was reliably longer than when the easy word came first (by 40 msec, and by 51 msec in the additional experiment). In other words, the hard names were also harder to access in sentence context. Still, the speakers could not capitalize on this noticeable difference in response speed by planning their utterances accordingly.

This result supports the notion of informational encapsulation. The conceptual system is ignorant about the accessibility of lexical items. Conceptual ordering decisions will not be reversed by such factors. The situation may, however, be slightly more complicated, as we will see shortly.

Is there feedback from the level of phonological encoding to that of grammatical encoding? Here three experimental results should be reported. The first finding, also from the Levelt-Maassen experiment, has to do with the choice of syntactic frame. Remember that conjoined or "common fate" movements (both figures rising together, or both falling) were preferably expressed by noun-phrase conjunction (e.g., *triangle and square went up*). There were, however, cases where such events were described by sentence conjunction (*triangle went up and square went up*). This occurred especially when both figures were difficult to name. Why did this happen? Levelt and Maassen conjectured that when the speaker had to retrieve two difficult word forms in close succession, he may not have succeeded in finding the second one in time to utter it when it was needed. The speaker may have started saying *triangle* ..., but then failed to come up in time with the phonetic plan for *square*. The speech need then "forced" him to give up constructing a noun-phrase conjunct, and to take a different route of grammatical encoding—one where the already-retrieved verb form (i.e., *went up*) could be uttered first. In other words, the speaker *revised* the grammatical encoding of the event. This, surely *is* a feedback explanation, and it should be supported by independent evidence. Such evidence exists. An on-the-fly revision of grammatical frame should take extra time. Therefore, the durations of these utterances were measured. And indeed, these critical utterances turned out to span more time (85 milliseconds more) than their controls. In other words, trouble in phonological encoding may affect grammatical encoding.

The issue was further tested in two experiments by Bock. In a first study, Bock (1986a) used her prime/recognition technique with pictures such as that in figure 7.2a above. However, this time the prime was not a spoken sentence but a spoken word. The word was either a semantic or a phonological prime. In the case of figure 7.2a, a semantic prime could be the word *worship*. It is a close associate of *church*. Bock conjectured that that prime might induce a picture description which would have *church* in a prominent grammatical role—i.e., something like the passive *The church is being struck by lightning*. Inversely, when the prime was *thunder*, its close associate *lightning* might end up in subject position, as in the active *The lightning is striking the church*. Such a finding would be very interesting in itself, because it is not so obvious that presenting the associate would affect the *conceptual* accessibility of the referent. It would, rather, directly affect the activation of the corresponding lemma. But the consequences should be the same if the analysis in the previous paragraph is correct: A lemma that is more highly activated, whether through saliency of the corresponding concept or through association with another activated lemma, will tend to end up in a more prominent grammatical function.

A phonological prime *sounds* like the target—e.g., *search* for *church* or *frightening* for *lightning*. And these kinds of primes form the critical test for the present issue. If phonological priming affects grammatical encoding, that is strong evidence for feedback from the level of phonological encoding to that of grammatical encoding. If, for instance, the prime *search* were to induce *church* to turn up in a particular grammatical function, this could only have been mediated by the phonological form of *church*.

Bock found that semantic primes did affect grammatical encoding. The primed element tended to turn up in subject position. However, there was no effect of phonological priming. In other words, no evidence was found for feedback from phonological to grammatical encoding.

But Bock (1987b) dedicated a second study to this issue. She applied the same priming/recognition method as in the previous research, but the experiments were different in three respects: First, in addition to scenes such as those presented in figure 7.2a, Bock presented pictures like those in figure 7.4. These were designed to release *conjunctive* descriptions, such as *The woman is carrying a lamp and a plant* and *The dog and the cat are under*

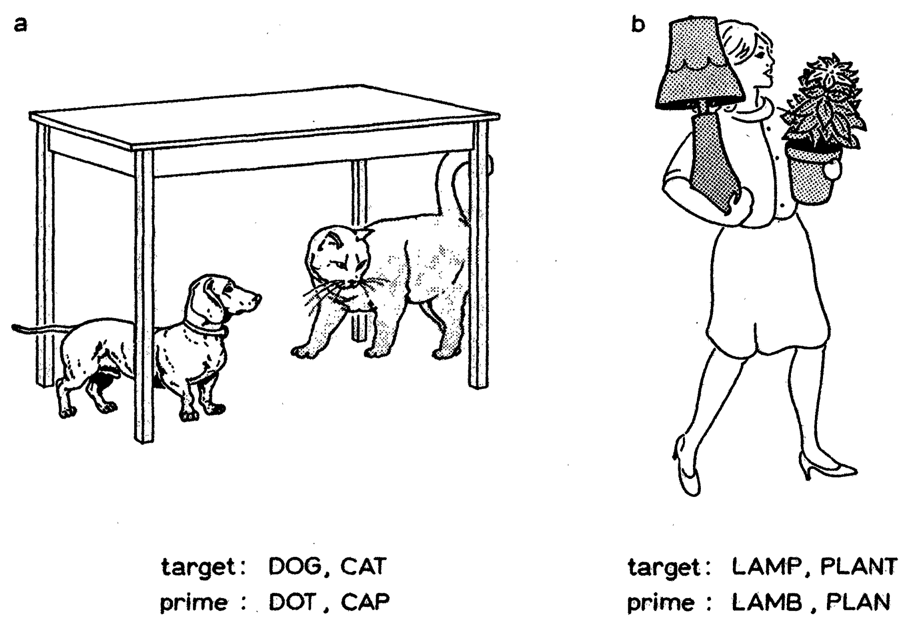


Figure 7.4
Examples of pictures types used by Bock (1987b) to elicit conjunctive descriptions.

the table. This is the kind of description Levelt and Maassen studies in their experiments. Second, the phonological primes were of a very strong kind. In all cases the word-initial part of the prime was identical to that of the target object. (In general, the phonological match was made as close as possible.) Examples of targets and primes are given in figure 7.4. Third, only phonological primes were used; there were no semantic primes.

This time Bock did obtain a significant phonological-priming effect. The effect, moreover, was one of *inhibition*: The primed target word tended to appear later in the sentence than the unprimed target word. For instance, when figure 7.4a was presented after the prime word *cap*, subjects tended to describe it as *The dog and the cat are under the table*, with the primed *cat* in secondary position. The effects were strongest for pictures like 7.2a. In 56.1 percent of these cases, sentences were constructed in which the primed word appeared in secondary position (i.e., not in subject position). The weakest effects were those for the conjunctive cases, as in figures 7.4a and 7.4b. In 52.9 percent of these cases the primed target word appeared in secondary position. But this was still significantly different from a random 50 percent result.

Bock also ran two control experiments. In the first one, subjects were not asked to describe the scene, but to just mention its two most prominent objects. The subjects' order-of-mention results showed the same kind and size of priming effects as had been obtained in the sentential descriptions. In the second control experiment, the subject did not have to repeat the prime word aloud, as in all previous experiments. Instead they were asked either to mouth it for themselves or to just listen to it. For both groups the results were essentially the same as in the main experiment.

What can we conclude from these excellent experiments? A first conclusion concerns the results with respect to scenes like that in figure 7.2a. These showed the strongest amount of (inhibitory) priming. The following explanation could be proposed: The phonological prime apparently inhibits the activation of the target word form. This, in turn, inhibits activation of the target lemma. This reduces the chance for that lemma to "claim" a prominent syntactic function (i.e., subject) during grammatical encoding. In other words, there is feedback from the phonological to the grammatical level of encoding. A slightly different account could bring this result in line with the feedback conjectured by Levelt and Maassen: Grammatical encoding initially proceeds without concern for potential trouble at the phonological level. But as soon as the first target name is about to be phonologically encoded, there will be a "speech need" situation if the name's phonological form is in an inhibited state due to the prime. This "trouble signal" is fed back to the level of grammatical encoding, and the syntactic frame is revised. Such revision took extra time in the Levelt-Maassen findings. Bock found a similar effect: There were dysfluencies in utterances, and these occurred mainly around the early target locations in the sentence. That is exactly the moment where syntactic revision should be taking place, if it occurs at all. On this account, the inhibition of the word form does *not* lead to inhibition of the corresponding lemma, but to a syntactic revision.

What are the theoretical implications of the results for conjunctive phrases, i.e., for the descriptions of scenes such as those in figure 7.4? It is obviously preferable to explain these in the same way as the previous ones. But the first kind of explanation above is not applicable. If indeed one lemma is inhibited by feedback from the phonological level, this will not have an effect in a situation where both lemmas end up in the same grammatical role. And that is the case in phrasal conjuncts. One cannot say anymore that the uninhibited lemma will be quicker to claim a more prominent grammatical function. For these scenes the only issue is the *order* of conjoining in the same grammatical function. That leaves us with the second account: syntactic revision in case of "speech need."

Though the revision account is the more economical one, covering both kinds of results in Bock's study as well as Levelt and Maassen's findings on feedback, one troublesome issue remains: Levelt and Maassen did *not* find an effect of lexical availability on ordering in conjunctive phrases, whereas Bock did. The difference, however, is more apparent than real. Levelt and Maassen found a nonsignificant ordering effect of 52.5 percent; Bock obtained a significant effect of 52.9 percent—a negligible difference. What remains to be dealt with is Levelt and Maassen's interpretation of their own test. Remember that they started from the assumption that the ordering decision is a conceptual one, and their hypothesis was that word-finding trouble should not lead to a revision of that decision. In retrospect, their experiments probably did not test that hypothesis. What they did test was whether phonological accessibility could affect syntactic encoding operations. These revised syntactic encoding operations would then *overrule* the conceptual input, but would not cause it to be *revised*. This interpretation is, moreover, in full agreement with the account of Bock's syntactic-biasing experiments given in section 7.3 above.

To sum up: The findings discussed in this section show the possibility of feedback from the phonological to the grammatical level of encoding. This feedback is likely to proceed as follows: Trouble at the level of access to word forms may induce a revision of syntactic frame. There are no findings that support on-line feedback at the next higher level of speech planning. The present evidence suggests that message preparation proceeds in full ignorance of the developing state of affairs at the levels of grammatical and phonological encoding.

Summary

Grammatical encoding takes a message as input and delivers a surface structure as output. It is likely that this process is highly automatic and nonintentional. A speaker will not, for every message, consider which of various grammatical alternatives would be most effective in reaching some communicative goal.

The first section of this chapter sketched a possible architecture for such automatic encoding: Kempen and Hoenkamp's (1987) Incremental Production Grammar. The algorithm has several attractive features as a model of the speaker: It is lexically driven; it generates "incrementally," from left to right, taking whatever message fragments the speaker makes available; it generates major constituents in parallel (which, as we saw, is a requirement if the fluency of speech is to be explained); it generates linguistically well-formed structures of an interesting, nontrivial variety; and so on. The model is, primarily, an existence proof. It shows the feasibility of constructing nonintentional grammatical-encoding algorithms that behave like real speakers in important respects.

The architecture of a processing system can sometimes become transparent in the ways it errs or fails. An effort was made to relate several types of speech error to joints in the Kempen-Hoenkamp architecture. Exchanges of words and of phrases could be attributed to errors in the delivery or the destination of procedural output. The algorithm could also predict the stranding of inflectional elements and of stress. These predictions met with varying but encouraging degrees of support from the error data.

Apart from destination errors, we distinguished errors of ordering. These should be attributed to the selection of possible but inappropriate ordering frames by categorial procedures in the algorithm. Such errors may indeed occur, but they certainly don't occur in large quantities. Rather more frequent are *shifts*—mostly slight mislocations of a word or a smaller segment. With Garrett, we attributed these to failures at the next level of processing, where phonological form is encoded.

The chapter then turned to various areas of experimental research in grammatical encoding. First, the studies of Ford and Holmes were mentioned. These studies suggest the existence of an encoding rhythm going from basic clause to basic clause—i.e., following the major semantic joints in the message rather than syntactic surface constituents. Second, the encoding of topical and highly accessible entities was discussed. It was shown that topical information is usually encoded early in the sentence, and preferably as grammatical subject. Also considered was how the various arguments in the thematic structure of the message would be distributed over grammatical functions, and over word-order positions. Grammatical encoding was called "congruent" if the relative importance of the various arguments and modifiers was reflected in the assignment of grammatical functions and constituent positions. Several studies were reviewed that showed that a systematic congruency mechanics exists, and that it is universal to an interesting degree. Third, some remarks were made on the generation of definiteness, pronominalization, and (de-)accentuation. These are anaphoric grammatical devices for linking the interpretation of the current utterance to what was said earlier or what is about to be said. They contribute to the coherence of discourse. Cross-linguistic evidence has shown both similarities and differences in the grammatical devices speakers use to establish coherence of reference. There is, in addition, syntactic coherence in the sense that an utterance repeats syntactic features of a previous utterance in the discourse. This phenomenon, which can be experimentally induced, shows that grammatical encoding procedures can become biased, independent of message-level input.

Finally, the issue of feedback was discussed. Can grammatical encoding affect message construction, and can phonological encoding affect grammatical encoding? The conclusions were fairly straightforward: There is no experimental support for the first kind of feedback. We can maintain the modular assumption that message construction proceeds without on-line feedback from the level of grammatical encoding. There is, however, clear evidence for the possibility of feedback from phonological to grammatical encoding. A minimal interpretation of this feedback is that it consists of revising a syntactic frame when trouble arises at the (next) level of phonological encoding.

Chapter 8

Phonetic Plans for Words and Connected Speech

The first stage of the formulating process, grammatical encoding, is followed by a second stage in which a representation of the utterance's form is generated. It takes successive fragments of surface structure as they become available as input; it produces, incrementally, the form specifications that the Articulator will have to realize, the speaker's phonetic plan. In going from a surface string of lemmas to a phonetic plan for connected speech, the speaker generates a variety of intermediary representations. Phonological encoding is not as simple as retrieving stored phonetic plans for successive words and concatenating them. Rather, the phonetic plan is a rhythmic (re-)syllabification of a string of segments. Each word's segments and basic rhythm are somehow stored in the form lexicon. When these word patterns are concatenated, new patterns arise. Segments may get lost or added, particularly at word boundaries. Syllables may be formed that cross word boundaries. Word accents are shifted to create a more regular rhythm for the larger string as a whole, and so on. Many of these operations serve to create more easily *pronounceable* patterns. And pronounceable is what the input to the Articulator should be. The present chapter reviews some of the form specifications involved in the generation of phonetic plans. How these target representations are built by the speaker will be the subject of chapters 9 and 10.

The review will be done in two steps. Section 8.1 is concerned with the form of words. Words, as we saw in chapter 6, have morphological and phonological structure. This structure is specified in the speaker's form lexicon, or can be composed from smaller lexical elements in the course of speaking. The process by which the speaker retrieves this form information from the lexicon and uses it to create a phonetic plan for the word suggests the existence of a multi-level organization of word-form properties. Section 8.2 deals with connected speech. The phonological encoding of surface structure often requires a phonetic plan that spans several words. Such