explanations for the rank frequency relation use only two or three parameters. The most they can hope to accomplish, therefore, is to provide a null hypothesis and to indicate in a qualitative way (perhaps) the kind of systems we are dealing with. They can tell us, for example, that any grammatical rule regulating word lengths must be regarded with considerable suspicion—in an English grammar, at least.

The complexity of the underlying linguistic process cannot be suppressed very far, however, and examples of nonrandom aspects are in good supply. For example, if we partition a random population on the basis of some independent criterion, the same probability distribution should apply to the partitions as to the parent population. If, for example, we partitioned according to whether the words were an odd or an even number of running words away from the beginning of the text or according to whether their initial letters were in the first or the last half of the alphabet, etc., we would expect the same rank-frequency relation to apply to the partitions as to the original population. There are, however, several ways to partition the parent population that look as though they ought to be independent but turn out in fact not to be. Thus, for example, Yule (1944) established that the same distribution does not apply when different categories (nouns, verbs, and adjectives) are taken separately; Miller, Newman, and Friedman (1958) showed a drastic difference between the distributions of content words (nouns, verbs, adjectives, adverbs) and of function words (everything else), and Miller (1951, p. 93) demonstrated that the distribution can be quite different if we consider only the words that occur immediately following a given word, such as *the* or *of*. There is nothing in our present parsimonious theories of the rank-frequency relation that could help us to explain these apparent deviations from randomness.

In an effort to achieve a more appropriate level of complexity in our descriptions of the user, therefore, we turn next to models that take account of the underlying structure of natural languages—models that, for lack of a better name, we shall refer to here as algebraic.

2. ALGEBRAIC MODELS

If the study of actual linguistic behavior is to proceed very far, it must clearly pay more than passing notice to the competence and knowledge of the performing organism. We have suggested that a generative grammar can give a useful and informative characterization of the competence of the speaker-hearer, one that captures many significant and deep-seated aspects of his knowledge of his own language. The question is, therefore, how does he put his knowledge to use in producing a desired sentence or in perceiving and interpreting the structure of presented utterances? How can we construct a model for the language user that incorporates a generative grammar as a fundamental component? This topic has received almost no study, so we can do little more than introduce a few speculations.

As we observed in the introduction to this chapter, models of linguistic performance can generally be interpreted interchangeably as depicting the behavior of either a speaker or a hearer. For concreteness, in the present sections we shall concentrate on the listener's task and frame our discussion largely in perceptual terms. This decision is, however, a matter of convenience, not of principle.

Unfortunately, the bulk of the experimental research on speech perception has involved the recognition of individual words spoken in isolation as part of a list (cf. Fletcher, 1953) and so is of little value to us in understanding the effects of grammatical structure on speech perception. That such effects exist is clear from the fact that the same words are easier to hear in sentences than in isolation (Miller, Heise, & Lichten, 1951; Miller, 1962a). How these effects are caused, however, is not at all clear.

Let us take as our starting point the sentence-recognizing device introduced briefly in Chapter 11, Sec. 6.4. Instead of a relatively passive process of acoustic analysis followed by identification and symbolic representation, we imagined (following Halle & Stevens, 1959, 1962) an active device that recognizes its input by discovering what must be done in order to generate a signal (in some possibly derived form) to match it. At the heart of this active device, of course, is a component M that contains rules for generating a matching signal. Associated with M would be components to analyze and (temporarily) to store the input, components that reflect various semantic and situational constraints suggested by the context of the sentence, a heuristic component that could make a good first guess, a component to make the comparison of the input and the internally generated signals, and perhaps others. On the basis of an initial guess, the device generates an internal signal according to the rules stored in M and tests its guess against the input signal. If the match is unsatisfactory, the discrepancy is used to make a better guess. In this manner the device proceeds to modify its own internal signal until the match is judged satisfactory or the input is dismissed as unintelligible. The program for generating the matching signal can be taken as the symbolic representation of the input.

If it is granted that such a sentence-recognizer can provide a plausible model for human speech perception, we can take it as our starting point and can proceed to try to specify it more precisely. In particular, the two parts of it that seem to perform the most important functions are the contextual component, which helps to generate a first guess, and the grammatical component M, which imposes the rules for generating the internal signal. We should begin by studying those two components. Even if it were feasible, a study of the ways contextual information can be stored and brought to bear would lead us far beyond the limits we have placed on this discussion. With respect to M, however, the task seems easier. The way the rules for synthesizing sentences might operate is, of course, very much in our present line of sight.

We are concerned with a finite device M in which are stored the rules of a generative grammar G. This device takes as its input a string x of symbols and attempts to understand it; that is to say, M tries to assign to x a certain structural description $F(x)$—or a set $\{F_1(x), \ldots, F_m(x)\}$ of syntactic descriptions in the case of a sentence x that is structurally ambiguous in m different ways. We shall not try to consider all of those real but obscure aspects of understanding that go beyond the assignment of syntactic structural descriptions to sentences, nor shall we consider the situational or contextual features that may determine which of a set of alternative structural descriptions is actually selected in a particular case. There is no point of principle underlying this limitation to syntax rather than to semantics and to single sentences rather than their linguistic and extra-linguistic contexts—it is simply an unfortunate consequence of limitations in our current knowledge and understanding. At present there is little that can be said, with much precision, about those further questions. [See Ziff (1960) and Katz & Fodor (1962) for discussion of the problems involved in the development of an adequate semantic theory and some of the ways in which they can be investigated].

The device M must contain, in addition to the rules of G, a certain amount of computing space, which may be utilized in various different ways, and it must be equipped to perform logical operations of various sorts. We require, in particular, that M assign a structural description $F_i(x)$ to x only if the generative grammar G stored in the memory of M assigns $F_i(x)$ to x as a possible structural description. We say that the device M *(partially) understands the sentence x in the manner of G* if the set $\{F_1(x), \ldots, F_m(x)\}$ of structural descriptions provided by M with input x is (included in) the set assigned to x by the generative grammar G. In particular, M does not accept as a sentence any string that is not generated by G. (This restriction can, of course, be softened by introducing degrees of grammaticalness, after the manner of Sec. 1.5, but we shall not burden the present discussion with that additional complication.) M is thus a finite transducer in the sense of Chapter 12, Sec. 1.5. It uses its information concerning the set of all strings in order to determine which of them are sentences of the language it understands and to understand sentences belonging to this language. This information, we assume, is represented in the form of rules of the generative grammar G stored in the memory of M.

Before continuing, we should like to say once more that it is perfectly possible that M will not contain enough computing space to allow it to understand all sentences in the manner of the device G whose instructions it stores. This is no more surprising than the fact that a person who knows the rules of arithmetic perfectly may not be able to perform many computations correctly in his head. One must be careful not to obscure the fundamental difference between, on the one hand, a device M storing the rules G but having enough computing space to understand in the manner of G only a certain proper subset L' of the set L of sentences generated by G and, on the other hand, a device M^* designed specifically to understand only the sentences of L' in the manner of G. The distinction is perfectly analogous to the distinction between a device F that contains the rules of arithmetic but has enough computing space to handle only a proper subset Σ' of the set Σ of arithmetical computations and a device F^* that is designed to compute only Σ'. Thus, although identical in their behavior to F^* and M^*, F and M can improve their behavior without additional instruction if given additional memory aids, but F^* and M^* must be redesigned to extend the class of cases that they can handle. It is clear that F and M, the devices that incorporate competence whether or not it is realized in performance, provide the only models of any psychological relevance, since only they can explain the transfer of learning that we know occurs when memory aids are in fact made available.

In particular, if the grammar G incorporated in M exceeds any finite automaton in generative capacity, then we know that M will not be able to understand all sentences in the manner of G. There would be little reason to expect, a priori, that the natural languages learned by humans should belong to the special family of sets that can be generated by one-sided linear grammars (cf. Defs. 6 and 7, Chapter 12, Sec. 4.1) or by nonself-embedding context-free grammars (cf. Proposition 58 and Theorem 33, Chapter 12, Sec. 4.6). In fact, they do not, as we have observed several times. Consequently, we know that a realistic model M for the perceiver will incorporate a grammar G that generates sentences that M cannot understand in the manner of G (without additional aids). This conclusion should occasion no surprise; it leads to none of the paradoxical consequences that have occasionally been suggested. There has been much confusion about this matter and we should like to reemphasize the fact that the conclusion we have reached is just what should have been expected.

We can construct a model for the listener who understands a presented sentence by specifying the stored grammar G, the organization of memory, and the operations performable by M. We determine a class of perceptual models by stating conditions that these specifications must meet. In

Sec. 2.1 we consider perceptual models that store rewriting systems. Then in Sec. 2.2 we discuss possible features of perceptual models that incorporate transformational grammars.

2.1 Models Incorporating Rewriting Systems

Let us suppose that we have a language L generated by a context-sensitive grammar G that assigns to each sentence of L a P-marker—a labeled tree or labeled bracketing—in the manner we have already considered. What can we say about the understanding of sentences by the speaker of L? For example, what can we say about the class of sentences of his language that this speaker will be able to understand at all? If we construct a finite perceptual device M that incorporates the rules of G in its memory, to what extent will M be able to understand sentences in the manner of G?

In part, we answered this question in Sec. 4.6 of Chapter 12. Roughly, the answer was the following. Suppose that we say that *the degree of self-embedding of the P-marker Q is m* if m is the largest integer meeting the following condition: there is, in the labeled tree that represents Q, a continuous path passing through $m + 1$ nodes $N_0, \ldots, N_m$, each with the same label, where each N_i $(i \geqslant 1)$ is fully self-embedded (with something to the left and something to the right) in the subtree dominated by N_{i-1}; that is to say, the terminal string of Q can be written in the form

$$xy_0y_1 \cdots y_{m-1}zv_{m-1} \cdots v_1v_0w, \tag{33}$$

where N_m dominates z, and for each $i < m$, N_i dominates

$$y_i \cdots y_{m-1}zv_{m-1} \cdots v_i, \tag{34}$$

and none of the strings $y_0, \ldots, y_{m-1}, v_0, \ldots, v_{m-1}$ is null. Thus, for example, in Fig. 5 the degree of self-embedding is two.

In Sec. 4.6 of Chapter 12 we presented a mechanical procedure Ψ that can be regarded as having the following effect: given a grammar G and an integer m, $\Psi(G, m)$ is a finite transducer M that takes a sentence x as input and gives as output a structural description $F(x)$ (which is, furthermore, a structural description assigned to x by G) wherever $F(x)$ has a degree of self-embedding of no more than m; that is to say, where m is a measure of the computing space available to a perceptual model M, which incorporates the grammar G, M will partially understand sentences in the manner of G just to the extent that the degree of self-embedding of their structural descriptions is not too great. As the amount of computing space available to the device M increases, M will understand more deeply embedded structures in the manner of G. For any given sentence x there is an m sufficiently large so that the device M with computing space determined by m [i.e., the device $\Psi(G, m)$] will be capable of understanding x in the manner of G; M does not have to be redesigned to extend its capacities in this way. Furthermore, this is the best result that can be achieved, since self-embedding is, as was proved in Chapter 12, precisely the property that distinguishes context-free languages from the regular languages that can be generated (accepted) by finite automata.

In Chapter 12 this result was stated only for a certain class K of context-free grammars. We pointed out that the class K contains a grammar for every context-free language and that it is a straightforward matter to drop many, if not all, of the restrictions that define K. Extension to context-sensitive grammars is another matter, however, and the problem of finding an optimal finite transducer that understands the sentences of G as well as possible, for any context-sensitive G, has not been investigated at all. Certain approaches to this question are suggested by the results of Matthews', discussed in Chapter 12, Sec. 4.2, on asymmetrical context-sensitive grammars and PDS automata, but these have not yet been pursued.

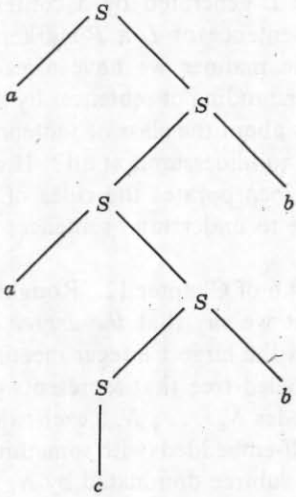

Fig. 5. Phrase marker with a degree of self-embedding equal to two.

These restrictions aside, the procedure Ψ of Sec. 4.6, Chapter 12, provides an optimal perceptual model (i.e., an optimal finite recognition routine) that incorporates a context-free grammar G. Given G, we can immediately construct such a device in a mechanical way, and we know that it will do as well as can be done by any device with bounded memory in understanding sentences in the manner of G. As the amount of memory increases, its capacity to understand sentences of G increases without limit. Only self-embedding beyond a certain degree causes it to fail when memory is fixed. We can, in fact, rephrase the construction so that the procedure Ψ determines a transducer $\Psi(G)$ which understands all sentences in the manner of G, where $\Psi(G)$ is a "single-pass" device with only pushdown storage, as shown in Sec. 4.2, Chapter 12.

Observe that the optimal perceptual model $M = \Psi(G, m)$, where m is fixed, may fail to understand sentences in the manner of G even when the language L generated by G might have been generated by a one-sided linear grammar (finite automaton). For example, the context-free grammar G that gives the structural description in Fig. 5 might be the following:

$$S \to aS, \qquad S \to Sb, \qquad S \to c. \tag{35}$$

(It is a straightforward matter to extend Ψ to deal with rules of the kind in Example 35.) The generated language is the set of all strings a^icb^j and is clearly a regular language. Nevertheless, with $m = 1$, $\Psi(G, m)$ will not be capable of understanding the sentence *aacbb* generated in Fig. 5 *in the manner of* G, since this derivation has a degree of self-embedding equal to two. The point is that although a finite automaton can be found to accept the sentences of this language it is not possible to find a finite device that understands all of its sentences in the manner of the particular generative process G represented in Example 35.

Observe also that the perceptual device $\Psi(G, m)$ is nondeterministic. As a perceptual model it has the following defect. Suppose that G assigns to x a structural description D with degree of self-embedding not exceeding m. Then, as we have indicated, the device $\Psi(G, m)$ will be capable of computing in such a way that it will map x into D, thus interpreting x in the manner of G. Being nondeterministic, however, it may also, given x, compute in such a way that it will fail to map x into a structural description at all. If $\Psi(G, m)$ fails to interpret x in the manner of G on a particular computation, we can conclude nothing about the status of x with respect to the grammar G, although if $\Psi(G, m)$ does map x into a structural description D we can conclude that G assigns D to x. We might investigate the problem of constructing a deterministic perceptual model that partially understands the output of a context-free grammar, or a model with nondeterminacy matching the ambiguity of the underlying grammar—that is, a model that may block on a computation with a particular string only if this string is either not generated by the grammar from which the model is constructed or is generated only by a derivation that is too deeply self-embedded for the device in question—but this matter has not yet been carefully investigated. It is clear, however, that such devices unlike $\Psi(G, m)$, would involve a restriction on the right-recursive elements in the structural descriptions (i.e., on right branchings). See, in this connection, the example on p. 473.

Self-embedding is the fundamental property that takes a system outside of the generative capacity of a finite device, and self-embedding will ultimately result from nesting of dependencies, since the nonterminal vocabulary is finite. However, the nesting of dependencies, even short of self-embedding, causes the number of states needed in the device $\Psi(G, m)$ to increase quite rapidly with the length of the input string that it is to understand. Consequently, we would expect that nested constructions should become difficult to understand even when they are, in principle, within the capacity of a finite device, since available memory (i.e., number of states) is clearly quite limited for real-time analytic operations, a fact to which we return in Sec. 2.2. Indeed, as we observed in Chapter 11 (cf. Example 11 in Sec. 3), nested structures even without self-embedding quickly become difficult or impossible to understand.

From these observations we are led to conclude that sentences of natural languages containing nested dependencies or self-embedding beyond a certain point should be impossible for (unaided) native speakers to understand. This is indeed the case, as we have already pointed out. There are many syntactic devices available in English—and in every other language that has been studied from this point of view—for the construction of sentences with nested dependencies. These devices, if permitted to operate freely, will quickly generate sentences that exceed the perceptual capacities (i.e., in this case, the short-term memory) of the native speakers of the language. This possibility causes no difficulties for communication, however. These sentences, being equally difficult for speaker and hearer, simply are not used, just as many other proliferations of syntactic devices that produce well-formed sentences will never actually be found.

There would be no reason to expect that these devices (which are, of course, continually used when nesting is kept within the bounds of memory restriction) should disappear as the language evolves; and, in fact, they do not disappear, as we have observed. It would be reasonable to expect, however, that a natural language might develop techniques to paraphrase complex nested sentences as sentences with either left-recursive or right-recursive elements, so that sentences of the same content could be produced with less strain on memory. That expectation, formulated by Yngve (1960, 1961) in a rather different way, to which we return, is well confirmed. Alongside such self-embedding English sentences as *if, whenever X then Y, then Z*, we can have the basically right-branching structure *Z if whenever X, then Y*, and so on in many other cases. In particular, many singulary grammatical transformations in English seem to be primarily stylistic; they convert one sentence into another with much the same content but with less self-embedding. Alongside the sentence *that the fact that he left was unfortunate is obvious*, which doubly embeds *S*, we have the more intelligible and primarily right-recursive structure *it is obvious that it was unfortunate that he left*. Similarly, we have a transformation that converts *the cover that the book that John has has* to *John's book's cover*, which is left-branching rather than self-embedding. (It should also be noted, however, that some of these so-called stylistic transformations can increase structural complexity, e.g., those that give "cleft-sentences"—from *I read the book that you told me about* we can form *it was the book that you told me about that I read*, etc.)

Now to recapitulate: from the fact that human memory is finite we can conclude only that some self-embedded structures should not be understandable; from the further assumption that memory is small, we can predict difficulties even with nested constructions. Although sentences are accepted (heard and spoken) in a single pass from left to right, we cannot conclude that there should be any left-right asymmetry in the understandable structures. Nor is there any evidence presently available for such asymmetry. We have little difficulty in understanding such right-branching constructions as *he watched the boy catch the ball that dropped from the tower near the lake* or such left-branching constructions as *all of the men whom I told you about who were exposed to radiation who worked half-time are still healthy, but the ones who worked full time are not* or *many more than half of the rather obviously much too easily solved problems were dropped last year*. Similarly, no conclusion can be drawn from our present knowledge of the distribution of left-recursive and right-recursive elements in language. Thus, in English, right-branching constructions predominate; in other languages—Japanese, Turkish—the opposite is the case. In fact, in every known language we find right-recursive, left-recursive, and self-embedding elements (and, furthermore, we find coordinate constructions that exceed the capacity of rewriting systems entirely, a fact to which we return directly).

We have so far made only the following assumptions about the model M for the user:

1. M is finite;
2. M accepts (or produces) sentences from left-to-right in a single pass;
3. M incorporates a context-free grammar as a representation of its competence in and knowledge of the language.

Of these, (3) is surely false, but the conclusions concerning recursive elements that we have drawn from it would undoubtedly remain true under a wide class of more general assumptions. Obviously, (1) is beyond question; (2) is an extremely weak assumption that also cannot be questioned, either for the speaker or hearer—note that many different kinds of internal organization of M are compatible with (2), for example, the assumption that M stores a finite string before deciding on the analysis of its first element or that M stores a finite number of alternative assumptions about the first element which are resolved only at an indefinitely later time.

If we add further assumptions beyond these three, we can derive additional conclusions about the ability of the device to produce or understand sentences in the manner of the incorporated grammar. Consider the two extreme assumptions:

4. *M* produces *P*-markers strictly "from the top down," or from trunk to branch, in the tree graph of the *P*-marker.
5. *M* produces *P*-markers strictly "from the bottom up," or from branch to trunk, in the tree graph of the *P*-marker.

In accordance with (4), the device *M* will interpret a rule $A \rightarrow \phi$ of the incorporated grammar as the instruction "rewrite A as ϕ"—that is to say, as the instruction that, in constructing a derivation, a line of the form $\psi_1 A \psi_2$ can be followed by the line $\psi_1 \phi \psi_2$. Assumption 5 requires the device *M* to interpret each rule $A \rightarrow \phi$ of the grammar as the instruction "replace ϕ by A"—that is to say, in constructing an inverted derivation with S as its last line and a terminal string as its first line, a line of the form $\psi_1 \phi \psi_2$ can be followed by the line $\psi_1 A \psi_2$.

From Assumption 4 we can conclude that only a bounded number of successive left-branchings can, in general, be tolerated by *M*. Thus suppose that *M* is based on a grammar containing the rule $S \rightarrow SA$. After n applications of this left-branching rule the memory of a device meeting Assumptions 2 and 4 (under the natural interpretation) would have to store n occurrences of A for later rewriting and would thus eventually have to violate Assumption 1. On the other hand, from Assumption 5 we can conclude that only a bounded number of successive right-branchings can in general be tolerated. For example, suppose the underlying grammar contains right-branching rules: $A \rightarrow cA$, $B \rightarrow cB$, $A \rightarrow a$, and $B \rightarrow b$. In this case the device will be presented with strings $c^n a$ or $c^n b$. Now, although Assumption 2 still calls for resolution from left to right, Assumption 5 implies that no node in the *P*-marker can be replaced until all that it dominates is known, so that resolution must be postponed until the final symbol in the string is received. Thus the device would have to store n occurrences of c for later rewriting and, again, Assumption 1 must eventually be violated. Left-branching causes no difficulty under Assumption 5, of course, just as right-branching causes no difficulty in the case of Assumption 4. Thus Assumptions 4 and 5 impose left-right asymmetries (in opposite ways) on the set of structures that can be accepted or produced by *M*. Observe that the devices $\Psi(G, m)$, given by the procedure Ψ of Chapter 12, Sec. 4.6, need not meet either of the restrictions in Assumption 4 or 5; in constructing a particular *P*-marker, they may move up or down or both ways indefinitely often, just as long as self-embedding is restricted.

Assumption 4 might be interpreted as a condition to be met by the speaker; Assumption 5, as a condition to be met by the hearer. (Of course, if we design a model of the speaker to meet Assumption 4 and a model of the hearer to meet Assumption 5 simultaneously, we will severely restrict the possibility of communication between them.) If Assumption 4 described the speaker, we would expect him to have difficulty with left-branching constructions; if Assumption 5 described the listener, we would expect him to have difficulty with right-branching constructions. Neither assumption seems particularly plausible. There is no reason to think that a speaker must always select his major phrase types before the minor subphrases or his word categories before his words (Assumption 4). Similarly, although a listener obviously receives terminal symbols and constructs phrase types, there is no reason to assume that decisions concerning minor phrase types must uniformly precede those concerning major structural features of the sentence. Assumptions 4 and 5 are but two of a large set of possible assumptions that might be considered in specifying models of the user more fully. Thus we might introduce an assumption that there is a bound on the length of the string that must be received before a construction can be uniquely identified by a left-to-right perceptual model—and so on, in many other ways.

There has been some discussion of hypotheses such as Assumptions 4 and 5. For example, Skinner's (1957) proposal that "verbal operant responses" to situations (e.g., the primary nouns, verbs, adjectives) form the raw materials of which sentences are constructed by higher level "autoclitic" responses (grammatical devices, ordering, selecting, etc.) might be loosely interpreted as a variant of Assumption 5, regarded as an assumption about the speaker. Yngve (1960, 1961) has proposed a variant of (4) as an assumption about the speaker; his proposal is explicitly directed toward our present topic and so demands a somewhat fuller discussion.

Yngve describes a process by which a device that contains a grammar rather similar to a context-free grammar produces derivations of utterances, always rewriting the leftmost nonterminal symbol in the last line of the already constructed derivation and postponing any nonterminal symbols to the right of it. Each postponed symbol, therefore, is a promise that must be remembered until the time comes to develop it; as the number of these promises grows, the load on memory also grows. Thus Yngve defines a measure of *depth* in terms of the number of postponed symbols, so that left-branching, self-embedding, and multiple-branching all contribute to depth, whereas right-branching does not. (Note that the depth of postponed symbols and the degree of embedding are quite distinct measures.) Yngve observes that a model so constructed for the speaker will be able with a limited memory to produce structures that do not exceed a certain depth. He offers the hypothesis that Assumption 4, so interpreted, is a correct characterization of the speaker and that natural languages have developed in such a way as to ease the speaker's task by limiting the necessity for left-branching.

The arguments in support of this hypothesis, however, seem inconclusive. It is difficult to see why any language should be designed for the ease of the speaker rather than the hearer, and Assumption 4 in any form seems totally unmotivated as a requirement for the hearer; on the contrary, the opposite assumption, as we have noted, seems the better motivated of the two. Nor does (4) seem to be a particularly plausible assumption concerning the speaker, for reasons we have already stated. It is possible, of course, to construct sentences that have a great depth and that are quite unintelligible, but they characteristically involve nesting or self-embedding and thus serve merely to show that the speaker and hearer have finite memories—that is to say, they support only the obvious and unquestionable Assumptions 1 and 2, not the additional Assumption 4. In order to support Yngve's hypothesis, we would have to find unintelligible sentences whose difficulty was attributable entirely to left-branching and multiple-branching. Such examples are not readily produced. In order to explain why multiple-branching, which contributes to the measure of depth, does not cause more difficulty, Yngve treats coordinate constructions (e.g., conjunctions) as right-branching, which does not contribute to the number of postponed symbols. But this is perfectly arbitrary; they could just as well be treated as left-branching. The only correct interpretation for such constructions is in terms of multiple-branching from a single node—this is exactly the formal feature that distinguishes true coordinate constructions, with no internal structure, from others. As we have observed in Chapter 11, Sec. 5, such constructions are beyond the limits of systems of rewriting rules altogether. Hence the relative ease with which such sentences as Examples 18 and 20 of Chapter 11 can be understood contradicts not only Assumption 4 but even the underlying Assumption 3, of which 4 is an elaboration.

In short, there seems to be little that we can say about the speaker and the hearer beyond the obvious fact that they are limited finite devices that relate sentences and structural descriptions and that they are subject to the constraint that time is linear. From this, all that we can conclude is that self-embedding (and, more generally, nesting of dependencies) should cause difficulty, as indeed it does. It is also not without interest that self-embedding seems to impose a greater burden than an equivalent amount of nesting without self-embedding. Further speculations are, at the present time, quite unsupported.

2.2 Models Incorporating Transformational Grammars

There are surprising limitations on the amount of short-term memory available for human data processing, although the amount of long-term memory is clearly great (cf. Miller, 1956). This fact suggests that it might be useful to look into the properties of a perceptual model M with two basic components, M_1 and M_2, operating as follows: M_1 contains a small, short-term memory. It performs computations on an input string x as it is received symbol by symbol and transmits the result of these computations to M_2. M_2 contains a large long-term memory in which is stored a generative grammar G; the task of M_2 is to determine the deeper structure of the input string x, using as its information the output transmitted to it by M_1. (Sentence-analyzing procedures of this sort have been investigated by Matthews, 1961.)

The details of the operation of M_2 would be complicated, of course; probably the best way to get an appreciation of the functions it would have to perform is to consider an example in some detail. Suppose, therefore, that a device M, so constructed, attempts to analyze such sentences as

John is easy to please. (36)

John is eager to please. (37)

To these, M_1 might assign preliminary analyses, as in Fig. 6, in which inessentials are omitted. Clearly, however, this is not the whole story. In order to account for the way in which we understand these sentences, it is necessary for the component M_2, accepting the analysis shown in Fig. 6 as input, to give as output structural descriptions that indicate that in

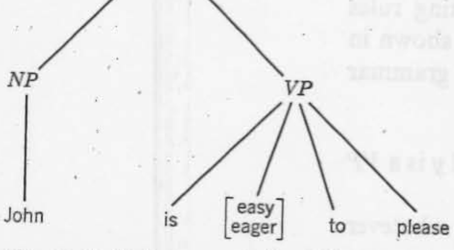

Fig. 6. Preliminary analysis of Sentences 36 and 37.

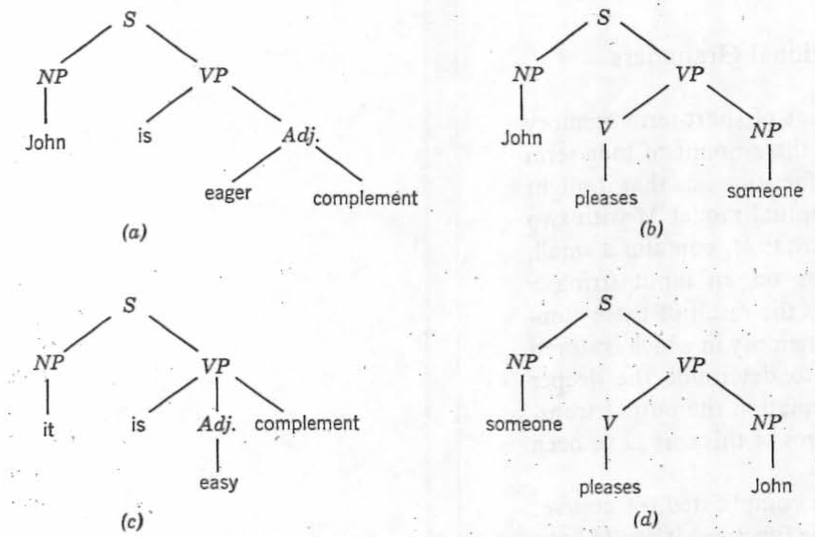

Fig. 7. Some *P*-markers that would be generated by the rewriting rules of the grammar and to which the transformation rules would apply.

Example 36 *John* is the direct object of *please*, whereas in Example 37 it is the logical subject of *please*.

Before we can attempt to provide a description of the device M_2 we must ask how structural information of this deeper kind can be represented. Clearly, it cannot be conveyed in the labeled tree (*P*-marker) associated with the sentence as it stands. No elaboration of the analysis shown in Fig. 6, with more elaborate subcategorization, etc., will remedy the fundamental inability of this form of representation to mirror grammatical relations properly. We are, of course, facing now precisely the kind of difficulty that was discussed in Chapter 11, Sec. 5, and that led to the development of a theory of transformational generative grammar. In a transformational grammar for English the rewriting rules would not be required to provide Examples 36 and 37 directly; the rewriting rules would be limited to the generation of such *P*-markers as those shown in Fig. 7 (where inessentials are again omitted). In addition, the grammar will contain such transformations as

T_1: replaces *complement* by "for x to y," where x is an *NP* and y is a *VP* in the already generated sentence xy;

T_2: deletes the second occurrence of two identical *NP*'s (with whatever is affixed to them);

T_3: deletes direct objects of certain verbs;

T_4: deletes "for someone" in certain contexts;
T_5: converts a string analyzable as

$$NP - \text{is} - Adj - (\text{for} - NP_1) - \text{to} - V - NP_2$$

to the corresponding string of the form

$$NP_2 - \text{is} - Adj - (\text{for} - NP_1) - \text{to} - V.$$

Each of these can be generalized and put in the form specified in Chapter 11. When appropriately generalized, they are each independently motivated by examples of many other kinds. Note, for example, the range of sentences that are similar in their underlying structural features to Examples 36 and 37; we have such sentences as *John is an easy person to please*, *John is a person who (it) is easy to please*, *this room is not easy to work in (to do decent work in)*, *he is easy to do business with*, *he is not easy to get information from*, *such claims are very easy to be fooled by*, and many others all of which are generated in essentially the same way.

Applying T_1 to the pair of structures in Figs. 7*c* and 7*d*, we derive the sentence *It is easy for someone to please John*, with its derived *P*-marker. Applying T_4 to this, we derive *It is easy to please John*, which is converted to Example 36 by T_5. Had we applied T_5 without T_4, we could have derived, for example, *John is easy for us to please* (with *we* chosen in place of *someone* in Fig. 7*d*—we leave unstated obvious obligatory rules). Applying T_1 to the pair of structures in Figs. 7*a* and 7*b*, we derive *John is eager for John to please someone*, which is converted by T_2 to *John is eager to please someone*. Had we applied T_3 to Fig. 7*b* before applying T_1, we would, in the same way, have derived Example 37.

At this point we should comment briefly on several features of such an analysis. Notice that *I am eager for you to please*, *you are eager for me to please*, etc., are all well-formed sentences; but *I am eager for me to please*, *you are eager for you to please*, etc., are impossible and are reduced to *I am eager to please*, *you are eager to please* obligatorily by T_2. This same transformation gives *I expected to come*, *you expected to come*, etc., from *I expected me to come*, *you expected you to come*, which are formed in the same way as *you expected me to come*, *I expected you to come*. Thus this grammar does actually regard *John* in Example 37 as identical with the deleted subject of *please*. Note, in fact, that in the sentence *John expected John to please*, in which T_2 has *not* applied, the two occurrences of *John* must have different reference. In Example 36, on the other hand, *John* is actually the direct object of *please*, assuming grammatical relations to be preserved under transformation (assuming, in other words, that the *P*-marker represented in Fig. 7*d* is part of the structural description of

Example 36). Note, incidentally, that T_5 does not produce such nonsentences as *John is easy to come*, since there is no *NP comes John*, though we have *John is eager to come* by T_1, T_2. T_5 would not apply to any sentence of the form

$$NP - \text{is} - \text{eager} - (\text{for} - NP_1) - \text{to} - V - NP_2$$

to give

$$NP_2 - \text{is} - \text{eager} - (\text{for} - NP_1) - \text{to} - V$$

(for example, *Bill is eager for us to meet* from *John is eager for us to meet Bill*; *these crooks are eager for us to vote out* from *John is eager for us to vote out these crooks*), since *eager complement*, but not *eager*, is an *Adj* (whereas, *easy*, but not *easy complement*, is an *Adj*). Supporting this analysis is the fact that the general rule that nominalizes sentences of the form *NP – is – Adj* (giving, for example, *John's cleverness* from *John is clever*), converts *John is eager (for us) to come* (which comes from Fig. 7*a* and *we come* by T_1) to *John's eagerness for us to come*; but it does not convert Example 36 to *John's easiness to please*. Furthermore, the general transformational process that converts phrases of the form

$$\text{the} - \text{Noun} - \text{who (which)} - \text{is} - Adj$$

to

$$\text{the} - Adj - \text{Noun}$$

(for example, *the man who is old* to *the old man*) does convert *a fellow who is easy to please* to *an easy fellow to please* (since *easy* is an *Adj*) but does not convert *a fellow who is eager to please* to *an eager fellow to please* (since *eager* is not, in this case, an *Adj*). In brief, when these rules are stated carefully, we find that a large variety of structures is generated by quite general, simple, and independently motivated rules, whereas other superficially similar structures are correctly excluded. It would not be possible to achieve the same degree of generalization and descriptive adequacy with a grammar that operates in the manner of a rewriting system, assigning just a single *P*-marker to a sentence as its structural description.

Returning now to our main theme, we see that the grammatical relations of *John* to *please* in Examples 36 and 37 are represented in the intuitively correct way in the structural descriptions provided by a transformational grammar. The structural description of Example 36 consists of the two underlying *P*-markers in Figs. 7*c* and 7*d* and the derived *P*-marker in Fig. 6 (as well as a record of the transformational history, i.e., T_1, T_4, T_5). The structural description of Example 36 consists of the underlying *P*-markers in Figs. 7*a* and 7*b* and the derived *P*-marker in Fig. 6 (along with the transformational history T_1, T_2, T_3). Thus the structural description of Example 36 contains the information that *John* in Example 36 is the object of *please* in the underlying *P*-marker of Fig. 7*d*; and the structural description of Example 37 contains the information that *John* in Example 37 is the subject of *please* in the underlying *P*-marker in Fig. 7*b*. Note that, when the appropriately generalized form of T_5 applies to *it is easy to do business with John* to yield *John is easy to do business with*, we again have in the underlying *P*-markers a correct account of the grammatical relations in the transform, although in this case the grammatical subject *John* is no longer the object of the verb of the complement, as it is in Example 3*b*. Notice also that it is the underlying *P*-markers, rather than the derived *P*-marker, that represent the semantically relevant information in this case. In this respect, these examples are quite typical of what is found in more extensive grammars.

These observations suggest that the transformational grammar be stored and utilized only by the component M_2 of the perceptual model. M_1 will take a sentence as input and give us as output a relatively superficial analysis of it (perhaps a derived *P*-marker such as that in Fig. 6). M_2 will utilize the full resources of the transformational grammar to provide a structural description, consisting of a set of *P*-markers and a transformational history, in which deeper grammatical relations and other structural information are represented. The output of $M = (M_1, M_2)$ will be the complete structural description assigned to the input sentence by the grammar that it stores; but the analysis that is provided by the initial, short-term memory component M_1 may be extremely limited.

If the memory limitations on M_1 are severe, we can expect to find that structurally complex sentences are beyond its analytic power even when they lack the property (i.e., repeated self-embedding) that takes them completely beyond the range of any finite device. It might be useful, therefore, to develop measures of various sorts to be correlated with understandability. One rough measure of structural complexity that we might use, along with degree of nesting and self-embedding, is the node-to-terminal-node ratio $N(Q)$ in the *P*-marker Q of the terminal string $t(Q)$. This number measures roughly the amount of computation per input symbol that must be performed by the listener. Hence an increase in $N(Q)$ should cause a correlated difficulty in interpreting $t(Q)$ for a real-time device with a small memory. Clearly $N(Q)$ grows as the amount of branching per node decreases. Thus $N(Q)$ is higher for a binary *P*-marker such as that shown in Fig. 8*a* than for the *P*-marker in Fig. 8*b* that represents a coordinate construction with the same number of terminals. Combined with our earlier speculations concerning the perceptual model M, this observation would lead us to suspect that $N(Q)$ should in general be higher for the derived *P*-marker that must be provided by the limited

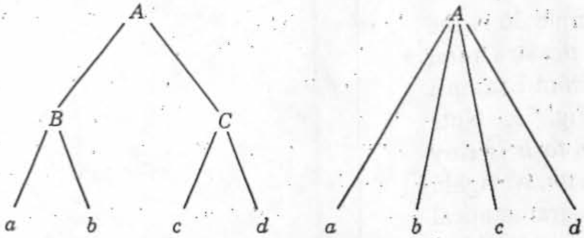

Fig. 8. Illustrating a measure of structural complexity. $N(Q)$ for the P-marker (a) is 7/4; for (b), $N(Q) = 5/4$.

component M_1 than it would be for underlying P-markers. In other words, the general effect of transformations should be to decrease the total amount of structure in the associated P-marker. This expectation is fully borne out. The underlying P-markers have limited, generally binary branching. But, as we have already observed in Chapter 11 (particularly p. 305), binary branching is not a general characteristic of the derived P-markers associated with actual sentences; in fact, the actual set of derived P-markers is beyond the generative capacity of rewriting systems altogether, since there is no bound on the amount of branching from a single node (that is to say, on the length of a coordinate construction).

The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring an appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved.

One source of psychological evidence concerns the grammatical transformation that negates an affirmative sentence. It is a well-established fact that people in concept-attainment experiments find it difficult to use negative instances (Smoke, 1933). Hovland and Weiss (1953) established that this difficulty persists even when the amount of information conveyed by the negative instances is carefully equated to the amount conveyed by positive instances. Moreover, Wason (1959, 1961) has shown that the grammatical difference between affirmative and negative English sentences causes more difficulty for subjects than the logical difference between true and false; that is to say, if people are asked to verify or to construct simple sentences (about whether digits in the range 2 to 9 are even or odd), they will take longer and make more errors on the true negative and false negative sentences than on the true affirmative and false affirmative sentences. Thus there is some reason to think that there may be a grammatical explanation for some of the difficulty we have in using negative information; moreover, this speculation has received some support from

Eifermann (1961), who found that negation in Hebrew has a somewhat different effect on thinking than it has in English.

A different approach can be illustrated by sentence-matching tests (Miller, 1962*b*). One study used a set of 18 elementary strings (for example, those formed by taking *Jane*, *Joe*, or *John* as the first constituent, *liked* or *warned* as the second, and *the old woman*, *the small boy*, or *the young man* as the last), along with the corresponding sets of sentences that could be formed from those by passive, negative, or passive-and-negative transformations. These sets were taken two at a time, and subjects were asked to match the sentences in one set with the corresponding sentences in the other. The rate at which they worked was recorded and from that it was possible to obtain an estimate of the time required to perform the necessary transformations. If we assume that these four types of sentence are coordinate and independently learned, then there is little reason to believe that finding correspondences between any two of them will necessarily be more difficult than between any other two. On the other hand, if we assume that the four types of sentence are related to one another by two grammatical transformations (and their inverses), then we would expect some of the tests to be much easier than others. The data supported a transformational position: the negative transformation was performed most rapidly, the more complicated passive transformation took slightly longer, and tests requiring both transformations (kernel to passive-negative or negative to passive) took as much time as the two single transformations did added together. For example, in order to perform the transformations necessary to match such pairs as *Jane didn't warn the small boy* and *The small boy was warned by Jane*, subjects required on the average more than three seconds, under the conditions of the test.

Still another way to explore these matters is to require subjects to memorize a set of sentences having various syntactic structures (J. Mehler, personal communication). Suppose, for example, that a person reads at a rapid but steady rate the following string of eight sentences formed by applying passive, negative, and interrogative transformations: *Has the train hit the car? The passenger hasn't been carried by the airplane. The photograph has been made by the boy. Hasn't the girl worn the jewel? The student hasn't written the essay. The typist has copied the paper. Hasn't the house been bought by the man? Has the discovery been made by the biologist?* When he finishes, he attempts to write down as many as he can recall. Then the list (in scrambled order) is read again, and again he tries to recall, and so on through a series of trials. Under those conditions many syntactic confusions occur, but most of them involve only a single transformational step. It is as if the person recoded the original sentences into something resembling a kernel string plus some correction terms for the transformations that indicate how to reconstruct the correct sentence when he is called on to recite. During recall he may remember the kernel, but become confused about which transformations to apply.

Preliminary evidence from these and similar studies seems to support the notion that kernel sentences play a central role, not only linguistically, but psychologically as well. It also seems likely that evidence bearing on the psychological reality of transformational grammar will come from careful studies of the genesis of language in infants, but we shall not attempt to survey that possibility here.

It should be obvious that the topics considered in this section have barely been opened for discussion. The problem can clearly profit from abstract study of various kinds of perceptual models that incorporate generative processes as a fundamental component. It would be instructive to study more carefully the kinds of structures that are actually found in natural languages and the formal features of those structures that make understanding and production of speech difficult. In this area the empirical study of language and the formal study of mathematical models may bear directly on questions of immediate psychological interest in what could turn out to be a highly fruitful and stimulating way.

3. TOWARD A THEORY OF COMPLICATED BEHAVIOR

It should by now be apparent that only a complicated organism can exploit the advantages of symbolic organization. Subjectively, we seem to grasp meanings as integrated wholes, yet it is not often that we can express a whole thought by a single sound or a single word. Before they can be communicated, ideas must be analyzed and represented by sequences of symbols. To map the simultaneous complexities of thought into a sequential flow of language requires an organism with considerable power and subtlety to symbolize and process information. These complexities make linguistic theory a difficult subject. But there is an extra reward to be gained from working it through. If we are able to understand something about the nature of human language, the same concepts and methods should help us to understand other kinds of complicated behavior as well.

Let us accept as an instance of complicated behavior any performance in which the behavioral sequence must be internally organized and guided by some hierarchical structure that plays the same role, more or less, as a *P*-marker plays in the organization of a grammatical sentence. It is not immediately obvious, of course, how we are to decide whether some particular nonlinguistic performance is complicated or simple; one natural criterion might be the ability to interrupt one part of the performance until some other part had been completed.

The necessity for analyzing a complex idea into its component parts has long been obvious. Less obvious, however, is the implication that any complicated activity obliges us to analyze and to postpone some parts while others are being performed. A task, X, say, is analyzed into the parts Y_1, Y_2, Y_3, which should, let us assume, be performed in that order. So Y_1 is singled out for attention while Y_2 and Y_3 are postponed. In order to accomplish Y_1, however, we find that we must analyze it into Z_1 and Z_2, and those in turn must be analyzed into still more detailed parts. This general situation can be expressed in various ways—by an outline or by a list structure (Newell, Shaw, & Simon, 1959) or by a tree graph similar to those used to summarize the structural description of individual sentences. While one part of a total enterprise is being accomplished, other parts may remain implicit and still largely unformulated. The ability to remember the postponed parts and to return to them in an appropriate order is necessarily reserved for organisms capable of complicated information processing. Thus the kind of theorizing we have been doing for sentences can easily be generalized to even larger units of behavior. Restricted-infinite automata in general, and PDS systems in particular, seem especially appropriate for the characterization of many different forms of complicated behavior.

The spectrum of complicated behavior extends from the simplest responses at one extreme to our most intricate symbolic processes at the other. In gross terms it is apparent that there is some scale of possibilities between these extremes, but exactly how we should measure it is a difficult problem. If we are willing to borrow from our linguistic analysis, there are several measures already available. We can list them briefly:

INFORMATION AND REDUNDANCY. The variety and stereotypy of the behavior sequences available to an organism are an obvious parameter to estimate in considering the complexity of its behavior (cf. Miller & Frick, 1949; Frick & Miller, 1951).

DEGREE OF SELF-EMBEDDING. This measure assumes a degree of complication that may seldom occur outside the realm of language and language-mediated behaviors. Self-embedding is of such great theoretical significance, however, that we should certainly look for occurrences of it in nonlinguistic contexts.

DEPTH OF POSTPONEMENT. This measure of memory load, proposed by Yngve, may be of particular importance in estimating a person's capacity to carry out complicated instructions or consciously to devise complicated plans for himself.

STRUCTURAL COMPLEXITY. The ratio of the total number of nodes in the hierarchy to the number of terminal nodes provides an estimate of complexity that, unlike the depth measure, is not asymmetrical toward the future.

TRANSFORMATIONAL COMPLEXITY. A hierarchical organization of behavior to meet some new situation may be constructed by transforming an organization previously developed in some more familiar situation. The number of transformations involved would provide an obvious measure of the complexity of the transfer from the old to the new situation.

These are some of the measures that we can adapt in analogy to the linguistic studies; no doubt many others of a similar nature could be developed.

Clearly, no one can look at a single instance of some performance and immediately assign values to it for any of those measures. As in the case of probability measures, repeated observations under many different conditions are required before a meaningful estimate is available.

Many psychologists, of course, prefer to avoid complicated behavior in their experimental studies; as long as there was no adequate way to cope with it, the experimentalist had little other alternative. Since about 1945, however, this situation has been changing rapidly. From mathematics and logic have come theoretical studies that are increasingly suggestive, and the development of high-speed digital computers has supplied a tool for exploring hypotheses that would have seemed fantastic only a generation ago. Today, for example, it is becoming increasingly common for experimental psychologists to phrase their theories in terms of a computer program for simulating behavior (cf. Chapter 7). Once a theory is expressed in that form, of course, it is perfectly reasonable to try to apply to it some of the indices of complexity.

Miller, Galanter, and Pribram (1960) have discussed the organization of complicated behavior in terms of a hierarchy of tote units. A *tote unit* consists of two parts: a *test* to see if some situation matches an internally generated criterion and an *operation* that is intended to reduce any differences between the external situation and some internal criterion. The criterion may derive from a model or hypothesis about what will be perceived or what would constitute a satisfactory state of affairs. The operations can either revise the criterion in the light of new evidence received or they can lead to actions that change the organism's internal and/or external environment. The test and its associated operations are actively linked in a feedback loop to permit iterated adjustments until the criterion is reached. A tote (test-operate-test-exit) unit is shown in the form of a flow-chart in Fig. 9. A hierarchy of tote units can be created by analyzing the operational phase into a sequence of tote units; then the operational phase of each is analyzed in turn. There should be no implication, however, that the hierarchy must be constructed exclusively from strategy to tactics or exclusively from tactics to strategy—both undoubtedly occur. An example of the kind of structures produced in this way is shown in the flowchart in Fig. 10.

These serial flowcharts are simply the finite automata we considered in Chapter 12, and it is convenient to replace them by oriented graphs (cf. Karp, 1960). Wherever an initial or terminal element or operation occurs in the flowchart, replace it by a node with one labeled arrow exiting from the node; wherever a test occurs, replace it by a node with two labeled exits. Next, replace every nonbranching sequence of arrows by a single arrow bearing a compound label. The graph corresponding to the flow-chart of Fig. 10 is shown in Fig. 11. From such oriented graphs as these it is a simple matter to read off the set of triples that define a finite automaton.

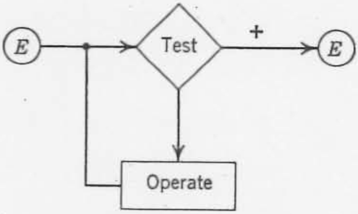

Fig. 9. A simple tote unit.

A tote hierarchy is just a general form of finite automaton in the sense of Chapter 12. We know from Theorem 2 of Chapter 12 that for any finite automaton there is an equivalent automaton that can be represented by a finite number of finite notations of the form $A_1(A_2, \ldots, A_m)^*A_{m+1}$, where the elements $A_2, \ldots, A_m$ can themselves be notations of the same form, and so on, until the full hierarchy is represented. For any finite state model that may be proposed, therefore, there is an equivalent model in terms of a (generalized) tote hierarchy.

Since a tote hierarchy is analogous to a program of instructions for a serial computer, it has been referred to as a *plan* that the system is trying to execute. Any postponed parts of the plan constitute the system's *intentions* at any given moment. Viewed in this way, therefore, the finite devices discussed in these chapters are clearly applicable to an even broader range of behavioral processes than language and communication. Some implications of this line of argument for nonlinguistic phenomena have been discussed informally by Miller, Galanter, and Pribram.

A central concern for this type of theory is to understand where new plans come from. Presumably, our richest source of new plans is our old plans, transformed to meet new situations. Although we know little about it, we must have ways to treat plans as objects that can be formed and transformed according to definite rules. The consideration of transformational grammars gives some indication of how we might combine

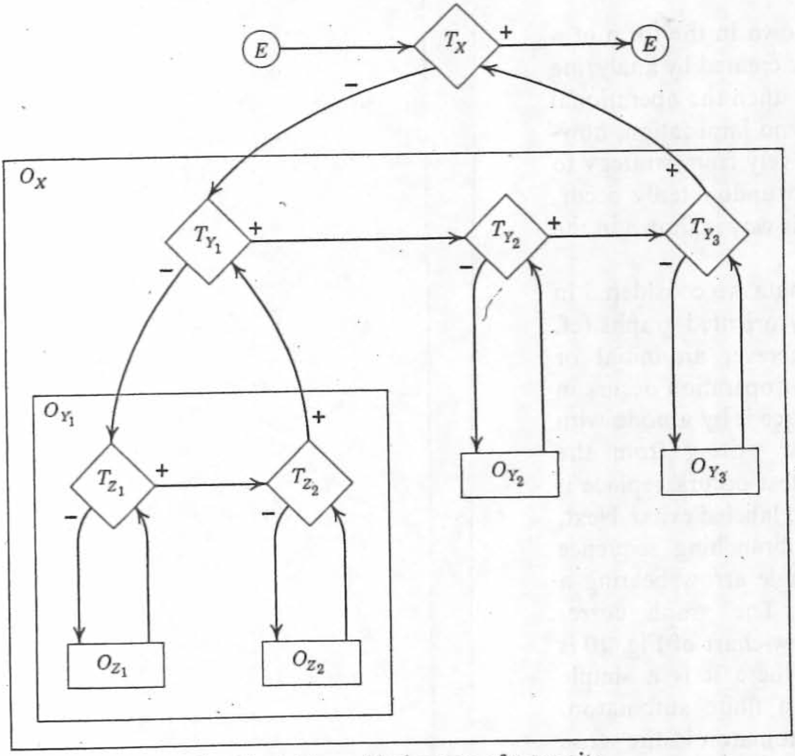

Fig. 10. A hierarchical system of tote units.

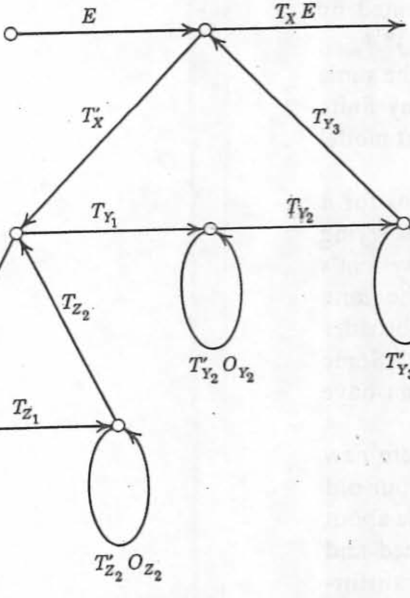

Fig. 11. Graph of flowchart in Fig. 10.

and rearrange plans, which are, of course, so closely analogous to *P*-markers. As in the case of grammatical transformations, the truly productive behavioral transformations are undoubtedly those that combine two or more simpler plans into one. These three chapters make it perfectly plain, however, how difficult it is to formulate a transformational system to achieve the twin goals of empirical adequacy and feasibility of abstract study.

When we ask about the source of our plans, however, we also raise the closely related question of what it might be that stands in the same relation to a plan as a grammar stands to a *P*-marker or as a programming language stands to a particular program. In what form are the rules stored whereby we construct, evaluate, and transform new plans? Probably there are many diverse sets of rules that govern our planning in different enterprises, and only patient observation and analysis of each behavioral system will enable us to describe the rules that govern them.

It is probably no accident that a theory of grammatical structure can be so readily and naturally generalized as a schema for theories of other kinds of complicated human behavior. An organism that is intricate and highly structured enough to perform the operations that we have seen to be involved in linguistic communication does not suddenly lose its intricacy and structure when it turns to nonlinguistic activities. In particular, such an organism can form verbal plans to guide many of its nonverbal acts. The verbal machinery turns out sentences—and, for civilized men, sentences have a compelling power to control both thought and action. Thus the present chapters, even though they have gone well beyond the usual bounds of psychology, raise issues that must be resolved eventually by any satisfactory psychological theory of complicated human behavior.

References

Attneave, F. *Applications of information theory to psychology.* New York: Holt-Dryden, 1959.

Burton, N. G., & Licklider, J. C. R. Long-range constraints in the statistical structure of printed English. *Amer. J. Psychol.*, 1955, **68**, 650–653.

Carnap, R., & Bar-Hillel, Y. *An outline of a theory of semantic information.* Res. Lab. Electronics, Cambridge: Mass. Inst. Tech. Tech. Rept. 247, 1952.

Chapanis, A. The reconstruction of abbreviated printed messages. *J. exp. Psychol.*, 1954, **48**, 496–510.

Cherry, C. *On human communication.* New York: Technology Press and Wiley, 1957.

Chomsky, N. *Logical structure of linguistic theory.* Microfilm. Mass. Inst. Tech. Libraries, 1955.

Condon, E. V. Statistics of vocabulary. *Science*, 1928, **67**, 300.

Cronbach, L. J. On the non-rational application of information measures in psychology. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 14–26.

Eifermann, R. R. Negation: a linguistic variable. *Acta Psychol.*, 1961, **18**, 258–273.

Estoup, J. B. *Gammes sténographique*. (4th ed.) Paris: 1916.

Fano, R. M. *The transmission of information*. Res. Lab. Electronics, Cambridge: Mass. Inst. Tech. Tech. Rept. 65, 1949.

Fano, R. M. *The transmission of information*. New York: Wiley, 1961.

Feinstein, A. *Foundations of information theory*. New York: McGraw-Hill, 1958.

Feller, W. *An introduction to probability theory and its applications*. (2nd ed.) New York: Wiley, 1957.

Fletcher, H. *Speech and hearing in communication*. (2nd ed.). New York: Van Nostrand, 1953.

Frick, F. C., & Miller, G. A. A statistical description of operant conditioning. *Amer. J. Psychol.*, 1951, **64**, 20–36.

Frick, F. C., & Sumby, W. H. Control tower language. *J. acoust. Soc. Amer.*, 1952, **24**, 595–597.

Fritz, E. L., & Grier, G. W., Jr. Pragmatic communications: A study of information flow in air traffic control. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 232–243.

Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

Gnedenko, B. V., & Kolmogorov, A. N. *Limit distributions for sums of independent random variables*. Translated by K. L. Chung. Cambridge, Mass.: Addison-Wesley, 1954.

Halle, M., & Stevens, K. N. Analysis by synthesis. In *Proc. Seminar on Speech Compression and Production*, AFCRC-TR-59-198, 1959.

Halle, M., & Stevens, K. N. Speech recognition: A model and a program for research. *IRE Trans. on Inform. Theory*, 1962, **IT-8**, 155–159.

Hardy, G. H., Littlewood, J. E., & Pólya, G. *Inequalities*. (2nd ed.). Cambridge: Cambridge Univer. Press, 1952.

Hartley, R. V. The transmission of information. *Bell System Tech. J.*, 1928, **17**, 535–550.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45**, 175–182.

Huffman, D. A. A method for the construction of minimum-redundancy codes. *Proc. IRE*, 1952, **40**, 1098–1101.

Karp, R. M. A note on the application of graph theory to digital computer programming. *Information and Control*, 1960, **3**, 179–190.

Katz, J., & Fodor, J. *The structure of a semantic theory*. To appear in *Language*. Reprinted in J. Katz & J. Fodor. *Readings in the philosophy of language*. New York: Prentice-Hall, 1963.

Khinchin, A. I. *Mathematical foundations of information theory*. Translated by R. A. Silverman and M. D. Friedman. New York: Dover, 1957.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Luce, R. D. (Ed.) *Developments in mathematical psychology*. Glencoe, Ill.: Free Press, 1960.

Mandelbrot, B. An informational theory of the structure of language based upon the theory of the statistical matching of messages and coding. In W. Jackson, (Ed.), *Proc. symp. on applications of communication theory*. London: Butterworth, 1953.

Mandelbrot, B. Linguistique statistique macroscopique. In L. Apostel, B. Mandelbrot, & A. Morf. *Logique, langage and théorie de l'information*. Paris: Universitaires de France, 1957. Pp. 1–78.

Mandelbrot, B. Les lois statistique macroscopiques du comportment. *Psychol. Française*, 1958, **3**, 237–249.

Mandelbrot, B. A note on a class of skew distribution functions: Analysis and critique of a paper by H. A. Simon. *Information and Control*, 1959, **2**, 90–99.

Mandelbrot, B. On the theory of word frequencies and on related Markovian models of discourse. In R. Jakobson (Ed.), *Structure of language in its mathematical aspect. Proc. 12th Symp. in App. Math.* Providence, R. I.: American Mathematical Society, 1961. Pp. 190–219.

Markov, A. A. Essai d'une recherche statistique sur le texte du roman "Eugene Onegin," *Bull acad. imper. Sci., St. Petersburg*, 1913, **7**.

Marschak, J. Remarks on the economics of information. In *Contributions to Scientific Research in Management*. Berkeley, Calif.: Univer. of California Press, 1960. Pp. 79–98.

Matthews, G. H. Analysis by synthesis of sentences of natural languages. In *Proc. 1st Int. Cong. on Machine Translation of Languages and Applied Language Analysis, 1961*. Teddington, England: National Physical Laboratory, (in press).

McMillan, B. The basic theorems of information theory. *Ann. math. Stat.*, 1953, **24**, 196–219.

Miller, G. A. *Language and communication*. New York: McGraw-Hill, 1951.

Miller, G. A. What is information measurement? *Amer. Psychologist*, 1953, **8**, 3–11.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Miller, G. A. Some effects of intermittent silence. *Amer. J. Psychol.*, 1957, **70**, 311–313.

Miller, G. A. Decision units in the perception of speech. *IRE Trans. Inform. Theory*, 1962, **IT-8**, No. 2, 81–83. (a)

Miller, G. A. Some psychological studies of grammar. *Amer. Psychologist*, 1962 **17**, 748–762. (b)

Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.

Miller, G. A., & Friedman, E. A. The reconstruction of mutilated English texts. *Information and Control*, 1957, **1**, 38–55.

Miller, G. A., Galanter, E., & Pribram, K. *Plans and the structure of behavior*. New York: Holt, 1960.

Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *J. exp. Psychol.*, 1951, **41**, 329–335.

Miller, G. A., & Newman, E. B. Tests of a statistical explanation of the rank-frequency relation for words in written English. *Amer. J. Psychol.*, 1958, **71**, 209–258.

Miller, G. A., Newman, E. B., & Friedman, E. A. Length-frequency statistics for written English. *Information and Control*, 1958, **1**, 370–398.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

Newell, A., Shaw, J. C., & Simon, H. A. Report on a general problem-solving program. In *Information Processing. Proc. International Conference on Information Processing, UNESCO*, Paris, June 1959. Pp. 256–264.

Newman, E. B. The pattern of vowels and consonants in various languages. *Amer. J. Psychol.*, 1951, **64**, 369–379.

Pareto, V. *Cours d'economie politique*. Paris: 1897.

Quastler, H. (Ed.). *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955.

Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.*, 1948, **27**, 379–423.

Shannon, C. E. Prediction and entropy of printed English. *Bell Syst. tech. J.*, 1951, **30**, 50–64.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Smoke, K. L. Negative instances in concept learning. *J. exp. Psychol.*, 1933, **16**, 583–588.

Somers, H. H. The measurement of grammatical constraints. *Language and Speech*, 1961, **4**, 150–156.

Thorndike, E. L., & Lorge, I. *The teacher's word book of* 30,000 *words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Toda, M. Information-receiving behavior in man. *Psychol. Rev.*, 1956, **63**, 204–212.

Wason, P. C. The processing of positive and negative information. *Quart. J. exp. Psychol.*, 1959, **11**, 92–107.

Wason, P. C. Response to affirmative and negative binary statements. *Brit. J. Psychol.*, 1961, **52**, 133–142.

Wiener, N. *Cybernetics*. New York: Wiley, 1948.

Willis, J. C. *Age and area*. Cambridge: Cambridge Univer. Press, 1922.

Yngve, V. H. A model and an hypothesis for language structure. *Proc. Am. Phil. Soc.*, 1960, **104**, 444–466.

Yngve, V. H. The depth hypothesis. In R. Jakobson (Ed.), *Structure of language and its mathematical aspect*. *Proc. 12th Symp. in App. Math.* Providence, R. I.: American Mathematical Society, 1961. Pp. 130–138.

Yule, G. U. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, FRS. *Phil. Trans. Roy. Soc.* (*London*), 1924, **B 213**, 21–87.

Yule, G. U. *The statistical study of literary vocabulary*. London: Cambridge Univer. Press, 1944.

Ziff, P. *Semantic analysis*. Ithaca: Cornell Univ. Press, 1960.

Zipf, G. K. *The psychobiology of language*. Boston: Houghton-Mifflin, 1935.

Zipf, G. K. *Human behavior and the principle of least effort*. Cambridge, Mass.: Addison-Wesley, 1949.