

I

Methodological Preliminaries

§ 1. *GENERATIVE GRAMMARS AS THEORIES OF LINGUISTIC COMPETENCE*

THIS study will touch on a variety of topics in syntactic theory and English syntax, a few in some detail, several quite superficially, and none exhaustively. It will be concerned with the syntactic component of a generative grammar, that is, with the rules that specify the well-formed strings of minimal syntactically functioning units (*formatives*) and assign structural information of various kinds both to these strings and to strings that deviate from well-formedness in certain respects.

The general framework within which this investigation will proceed has been presented in many places, and some familiarity with the theoretical and descriptive studies listed in the bibliography is presupposed. In this chapter, I shall survey briefly some of the main background assumptions, making no serious attempt here to justify them but only to sketch them clearly.

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. This seems to me to have been the position of the founders of modern general linguistics, and no cogent reason for

modifying it has been offered. To study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one. In this respect, study of language is no different from empirical investigation of other complex phenomena.

We thus make a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior.¹ Observed use of language or hypothesized dispositions to respond, habits, and so on, may provide evidence as to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline. The distinction I am noting here is related to the *langue-parole* distinction of Saussure; but it is necessary to reject his concept of *langue* as merely a systematic inventory of items and to return rather to the Humboldtian conception of underlying competence as a system of generative processes. For discussion, see Chomsky (1964).

A grammar of a language purports to be a description of the ideal speaker-hearer's intrinsic competence. If the grammar is, furthermore, perfectly explicit—in other words, if it does not rely on the intelligence of the understanding reader but rather provides an explicit analysis of his contribution—we may (somewhat redundantly) call it a *generative grammar*.

A fully adequate grammar must assign to each of an infinite range of sentences a structural description indicating how this

sentence is understood by the ideal speaker-hearer. This is the traditional problem of descriptive linguistics, and traditional grammars give a wealth of information concerning structural descriptions of sentences. However, valuable as they obviously are, traditional grammars are deficient in that they leave unexpressed many of the basic regularities of the language with which they are concerned. This fact is particularly clear on the level of syntax, where no traditional or structuralist grammar goes beyond classification of particular examples to the stage of formulation of generative rules on any significant scale. An analysis of the best existing grammars will quickly reveal that this is a defect of principle, not just a matter of empirical detail or logical preciseness. Nevertheless, it seems obvious that the attempt to explore this largely uncharted territory can most profitably begin with a study of the kind of structural information presented by traditional grammars and the kind of linguistic processes that have been exhibited, however informally, in these grammars.²

The limitations of traditional and structuralist grammars should be clearly appreciated. Although such grammars may contain full and explicit lists of exceptions and irregularities, they provide only examples and hints concerning the regular and productive syntactic processes. Traditional linguistic theory was not unaware of this fact. For example, James Beattie (1788) remarks that

Languages, therefore, resemble men in this respect, that, though each has peculiarities, whereby it is distinguished from every other, yet all have certain qualities in common. The peculiarities of individual tongues are explained in their respective grammars and dictionaries. Those things, that all languages have in common, or that are necessary to every language, are treated of in a science, which some have called *Universal* or *Philosophical* grammar.

Somewhat earlier, Du Marsais defines universal and particular grammar in the following way (1729; quoted in Sahlin, 1928, pp. 29-30):

Il y a dans la grammaire des observations qui conviennent à toutes les langues; ces observations forment ce qu'on appelle la grammaire

générale: telles sont les remarques que l'on a faites sur les sons articulés, sur les lettres qui sont les signes de ces sons; sur la nature des mots, et sur les différentes manières dont ils doivent être ou arrangés ou terminés pour faire un sens. Outre ces observations générales, il y en a qui ne sont propres qu'à une langue particulière; et c'est ce qui forme les grammaires particulières de chaque langue.

Within traditional linguistic theory, furthermore, it was clearly understood that one of the qualities that all languages have in common is their "creative" aspect. Thus an essential property of language is that it provides the means for expressing indefinitely many thoughts and for reacting appropriately in an indefinite range of new situations (for references, cf. Chomsky, 1964, forthcoming). The grammar of a particular language, then, is to be supplemented by a universal grammar that accommodates the creative aspect of language use and expresses the deep-seated regularities which, being universal, are omitted from the grammar itself. Therefore it is quite proper for a grammar to discuss only exceptions and irregularities in any detail. It is only when supplemented by a universal grammar that the grammar of a language provides a full account of the speaker-hearer's competence.

Modern linguistics, however, has not explicitly recognized the necessity for supplementing a "particular grammar" of a language by a universal grammar if it is to achieve descriptive adequacy. It has, in fact, characteristically rejected the study of universal grammar as misguided; and, as noted before, it has not attempted to deal with the creative aspect of language use. It thus suggests no way to overcome the fundamental descriptive inadequacy of structuralist grammars.

Another reason for the failure of traditional grammars, particular or universal, to attempt a precise statement of regular processes of sentence formation and sentence interpretation lay in the widely held belief that there is a "natural order of thoughts" that is mirrored by the order of words. Hence, the rules of sentence formation do not really belong to grammar but to some other subject in which the "order of thoughts" is studied. Thus in the *Grammaire générale et raisonnée* (Lancelot

et al., 1660) it is asserted that, aside from figurative speech, the sequence of words follows an "ordre naturel," which conforms "à l'expression naturelle de nos pensées." Consequently, few grammatical rules need be formulated beyond the rules of ellipsis, inversion, and so on, which determine the figurative use of language. The same view appears in many forms and variants. To mention just one additional example, in an interesting essay devoted largely to the question of how the simultaneous and sequential array of ideas is reflected in the order of words, Diderot concludes that French is unique among languages in the degree to which the order of words corresponds to the natural order of thoughts and ideas (Diderot, 1751). Thus "quel que soit l'ordre des termes dans une langue ancienne ou moderne, l'esprit de l'écrivain a suivi l'ordre didactique de la syntaxe française" (p. 390); "Nous disons les choses en français, comme l'esprit est forcé de les considérer en quelque langue qu'on écrive" (p. 371). With admirable consistency he goes on to conclude that "notre langue *pédestre* a sur les autres l'avantage de l'utile sur l'agréable" (p. 372); thus French is appropriate for the sciences, whereas Greek, Latin, Italian, and English "sont plus avantageuses pour les lettres." Moreover,

le bon sens choisirait la langue française; mais . . . l'imagination et les passions donneront la préférence aux langues anciennes et à celles de nos voisins . . . il faut parler français dans la société et dans les écoles de philosophie; et grec, latin, anglais, dans les chaires et sur les théâtres; . . . notre langue sera celle de la vérité, si jamais elle revient sur la terre; et . . . la grecque, la latine et les autres seront les langues de la fable et du mensonge. Le français est fait pour instruire, éclairer et convaincre; le grec, le latin, l'italien, l'anglais, pour persuader, émouvoir et tromper: parlez grec, latin, italien au peuple; mais parlez français au sage. (pp. 371-372)

In any event, insofar as the order of words is determined by factors independent of language, it is not necessary to describe it in a particular or universal grammar, and we therefore have principled grounds for excluding an explicit formulation of syntactic processes from grammar. It is worth noting that this naïve view of language structure persists to modern times in

various forms, for example, in Saussure's image of a sequence of expressions corresponding to an amorphous sequence of concepts or in the common characterization of language use as merely a matter of use of words and phrases (for example, Ryle, 1953).

But the fundamental reason for this inadequacy of traditional grammars is a more technical one. Although it was well understood that linguistic processes are in some sense "creative," the technical devices for expressing a system of recursive processes were simply not available until much more recently. In fact, a real understanding of how a language can (in Humboldt's words) "make infinite use of finite means" has developed only within the last thirty years, in the course of studies in the foundations of mathematics. Now that these insights are readily available it is possible to return to the problems that were raised, but not solved, in traditional linguistic theory, and to attempt an explicit formulation of the "creative" processes of language. There is, in short, no longer a technical barrier to the full-scale study of generative grammars.

Returning to the main theme, by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences. Obviously, every speaker of a language has mastered and internalized a generative grammar that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he can become aware of them, or that his statements about his intuitive knowledge of the language are necessarily accurate. Any interesting generative grammar will be dealing, for the most part, with mental processes that are far beyond the level of actual or even potential consciousness; furthermore, it is quite apparent that a speaker's reports and viewpoints about his behavior and his competence may be in error. Thus a generative grammar attempts to specify what the speaker actually knows, not what he may report about his knowledge. Similarly, a theory of visual perception would attempt to account for what a person actually sees and the mechanisms that determine this rather than his statements about what he sees and why, though these state-

ments may provide useful, in fact, compelling evidence for such a theory.

To avoid what has been a continuing misunderstanding, it is perhaps worth while to reiterate that a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker-hearer. When we speak of a grammar as generating a sentence with a certain structural description, we mean simply that the grammar assigns this structural description to the sentence. When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient way, to construct such a derivation. These questions belong to the theory of language use—the theory of performance. No doubt, a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of the language; but this generative grammar does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech production. For various attempts to clarify this point, see Chomsky (1957), Gleason (1961), Miller and Chomsky (1963), and many other publications.

Confusion over this matter has been sufficiently persistent to suggest that a terminological change might be in order. Nevertheless, I think that the term "generative grammar" is completely appropriate, and have therefore continued to use it. The term "generate" is familiar in the sense intended here in logic, particularly in Post's theory of combinatorial systems. Furthermore, "generate" seems to be the most appropriate translation for Humboldt's term *erzeugen*, which he frequently uses, it seems, in essentially the sense here intended. Since this use of the term "generate" is well established both in logic and in the tradition of linguistic theory, I can see no reason for a revision of terminology.

§ 2. TOWARD A THEORY OF PERFORMANCE

There seems to be little reason to question the traditional view that investigation of performance will proceed only so far as understanding of underlying competence permits. Furthermore, recent work on performance seems to give new support to this assumption. To my knowledge, the only concrete results that have been achieved and the only clear suggestions that have been put forth concerning the theory of performance, outside of phonetics, have come from studies of performance models that incorporate generative grammars of specific kinds — that is, from studies that have been based on assumptions about underlying competence.³ In particular, there are some suggestive observations concerning limitations on performance imposed by organization of memory and bounds on memory, and concerning the exploitation of grammatical devices to form deviant sentences of various types. The latter question is one to which we shall return in Chapters 2 and 4. To clarify further the distinction between competence and performance, it may be useful to summarize briefly some of the suggestions and results that have appeared in the last few years in the study of performance models with limitations of memory, time, and access.

For the purposes of this discussion, let us use the term "acceptable" to refer to utterances that are perfectly natural and immediately comprehensible without paper-and-pencil analysis, and in no way bizarre or outlandish. Obviously, acceptability will be a matter of degree, along various dimensions. One could go on to propose various operational tests to specify the notion more precisely (for example, rapidity, correctness, and uniformity of recall and recognition, normalcy of intonation).⁴ For present purposes, it is unnecessary to delimit it more carefully. To illustrate, the sentences of (1) are somewhat more acceptable, in the intended sense, than those of (2):

- (1) (i) I called up the man who wrote the book that you told me about
 (ii) quite a few of the students who you met who come from New York are friends of mine

- (iii) John, Bill, Tom, and several of their friends visited us last night
 (2) (i) I called the man who wrote the book that you told me about up
 (ii) the man who the boy who the students recognized pointed out is a friend of mine

The more acceptable sentences are those that are more likely to be produced, more easily understood, less clumsy, and in some sense more natural.⁵ The unacceptable sentences one would tend to avoid and replace by more acceptable variants, wherever possible, in actual discourse.

The notion "acceptable" is not to be confused with "grammatical." Acceptability is a concept that belongs to the study of performance, whereas grammaticalness belongs to the study of competence. The sentences of (2) are low on the scale of acceptability but high on the scale of grammaticalness, in the technical sense of this term. That is, the generative rules of the language assign an interpretation to them in exactly the way in which they assign an interpretation to the somewhat more acceptable sentences of (1). Like acceptability, grammaticalness is, no doubt, a matter of degree (cf. Chomsky, 1955, 1957, 1961), but the scales of grammaticalness and acceptability do not coincide. Grammaticalness is only one of many factors that interact to determine acceptability. Correspondingly, although one might propose various operational tests for acceptability, it is unlikely that a necessary and sufficient operational criterion might be invented for the much more abstract and far more important notion of grammaticalness. The unacceptable grammatical sentences often cannot be used, for reasons having to do, not with grammar, but rather with memory limitations, intonational and stylistic factors, "iconic" elements of discourse (for example, a tendency to place logical subject and object early rather than late; cf. note 32, Chapter 2, and note 9, Chapter 3), and so on. Note that it would be quite impossible to characterize the unacceptable sentences in grammatical terms. For example, we cannot formulate particular rules of the grammar in such a way as

to exclude them. Nor, obviously, can we exclude them by limiting the number of reapplications of grammatical rules in the generation of a sentence, since unacceptability can just as well arise from application of distinct rules, each being applied only once. In fact, it is clear that we can characterize unacceptable sentences only in terms of some "global" property of derivations and the structures they define — a property that is attributable, not to a particular rule, but rather to the way in which the rules interrelate in a derivation.

This observation suggests that the study of performance could profitably begin with an investigation of the acceptability of the simplest formal structures in grammatical sentences. The most obvious formal property of utterances is their bracketing into constituents of various types, that is, the "tree structure" associated with them. Among such structures we can distinguish various kinds — for example, those to which we give the following conventional technical names, for the purposes of this discussion:

- (3) (i) nested constructions
- (ii) self-embedded constructions
- (iii) multiple-branching constructions
- (iv) left-branching constructions
- (v) right-branching constructions

The phrases *A* and *B* form a nested construction if *A* falls totally within *B*, with some nonnull element to its left within *B* and some nonnull element to its right within *B*. Thus the phrase "the man who wrote the book that you told me about" is nested in the phrase "called the man who wrote the book that you told me about up," in (2i). The phrase *A* is self-embedded in *B* if *A* is nested in *B* and, furthermore, *A* is a phrase of the same type as *B*. Thus "who the students recognized" is self-embedded in "who the boy who the students recognized pointed out," in (2ii), since both are relative clauses. Thus nesting has to do with bracketing, and self-embedding with labeling of brackets as well. A multiple-branching construction is one with no internal structure. In (1iii), the Subject Noun Phrase is multiple-branch-

ing, since "John," "Bill," "Tom," and "several of their friends" are its immediate constituents, and have no further association among themselves. In terms of bracketing, a multiple-branching construction has the form $[[A][B]\cdots[M]]$. A left-branching structure is of the form $[[[\cdots]\cdots]\cdots]$ — for example, in English, such indefinitely iterable structures as $[[[[[John's\ brother]'s\ father]'s\ uncle]$ or $[[[the\ man\ who\ you\ met]\ from\ Boston]\ who\ was\ on\ the\ train]$, or (1ii), which combines several kinds of left-branching. Right-branching structures are those with the opposite property — for example, the Direct-Object of (1i) or $[this\ is\ [the\ cat\ that\ caught\ [the\ rat\ that\ stole\ the\ cheese]]]$.

The effect of these superficial aspects of sentence structure on performance has been a topic of study since almost the very inception of recent work on generative grammar, and there are some suggestive observations concerning their role in determining acceptability (that is, their role in limiting performance). Summarizing this work briefly, the following observations seem plausible:

- (4) (i) repeated nesting contributes to unacceptability
- (ii) self-embedding contributes still more radically to unacceptability
- (iii) multiple-branching constructions are optimal in acceptability
- (iv) nesting of a long and complex element reduces acceptability
- (v) there are no clear examples of unacceptability involving only left-branching or only right-branching, although these constructions are unnatural in other ways — thus, for example, in reading the right-branching construction "this is the cat that caught the rat that stole the cheese," the intonation breaks are ordinarily inserted in the wrong places (that is, after "cat" and "rat," instead of where the main brackets appear)

In some measure, these phenomena are easily explained. Thus it is known (cf. Chomsky, 1959a; and for discussion, Chomsky, 1961, and Miller and Chomsky, 1963) that an optimal perceptual

device, even with a bounded memory, can accept unbounded left-branching and right-branching structures, though nested (hence ultimately self-embedded) structures go beyond its memory capacity. Thus case (4i) is simply a consequence of finiteness of memory, and the unacceptability of such examples as (2ii) raises no problem.

If (4ii) is correct,⁶ then we have evidence for a conclusion about organization of memory that goes beyond the triviality that it must be finite in size. An optimal finite perceptual device of the type discussed in Chomsky (1959a) need have no more difficulty with self-embedding than with other kinds of nesting (see Bar-Hillel, Kasher, and Shamir, 1963, for a discussion of this point). To account for the greater unacceptability of self-embedding (assuming this to be a fact), we must add other conditions on the perceptual device beyond mere limitation of memory. We might assume, for example, that the perceptual device has a stock of analytic procedures available to it, one corresponding to each kind of phrase, and that it is organized in such a way that it is unable (or finds it difficult) to utilize a procedure φ while it is in the course of executing φ . This is not a necessary feature of a perceptual model, but it is a rather plausible one, and it would account for (4ii). See, in this connection, Miller and Isard (1964).

The high acceptability of multiple-branching, as in case (4iii), is easily explained on the rather plausible assumption that the ratio of number of phrases to number of formatives (the node-to-terminal node ratio, in a tree-diagram of a sentence) is a rough measure of the amount of computation that has to be performed in analysis. Thus multiple coordination would be the simplest kind of construction for an analytic device — it would impose the least strain on memory.⁷ For discussion, see Miller and Chomsky (1963).

Case (4iv) suggests decay of memory, perhaps, but raises un-solved problems (see Chomsky, 1961, note 19).

Case (4v) follows from the result about optimal perceptual models mentioned earlier. But it is unclear why left- and right-branching structures should become unnatural after a certain point, if they actually do.⁸

One might ask whether attention to less superficial aspects of grammatical structure than those of (3) could lead to somewhat deeper conclusions about performance models. This seems entirely possible. For example, in Miller and Chomsky (1963) some syntactic and perceptual considerations are adduced in support of a suggestion (which is, to be sure, highly speculative) as to the somewhat more detailed organization of a perceptual device. In general, it seems that the study of performance models incorporating generative grammars may be a fruitful study; furthermore, it is difficult to imagine any other basis on which a theory of performance might develop.

There has been a fair amount of criticism of work in generative grammar on the grounds that it slights study of performance in favor of study of underlying competence. The facts, however, seem to be that the only studies of performance, outside of phonetics (but see note 3), are those carried out as a by-product of work in generative grammar. In particular, the study of memory limitations just summarized and the study of deviation from rules, as a stylistic device, to which we return in Chapters 2 and 4, have developed in this way. Furthermore, it seems that these lines of investigation can provide some insight into performance. Consequently, this criticism is unwarranted, and, furthermore, completely misdirected. It is the descriptivist limitation-in-principle to classification and organization of data, to "extracting patterns" from a corpus of observed speech, to describing "speech habits" or "habit structures," insofar as these may exist, etc., that precludes the development of a theory of actual performance.

§ 3. THE ORGANIZATION OF A GENERATIVE GRAMMAR

Returning now to the question of competence and the generative grammars that purport to describe it, we stress again that knowledge of a language involves the implicit ability to understand indefinitely many sentences.⁹ Hence, a generative grammar must be a system of rules that can iterate to generate an in-

definitely large number of structures. This system of rules can be analyzed into the three major components of a generative grammar: the syntactic, phonological, and semantic components.¹⁰

The syntactic component specifies an infinite set of abstract formal objects, each of which incorporates all information relevant to a single interpretation of a particular sentence.¹¹ Since I shall be concerned here only with the syntactic component, I shall use the term "sentence" to refer to strings of formatives rather than to strings of phones. It will be recalled that a string of formatives specifies a string of phones uniquely (up to free variation), but not conversely.

The phonological component of a grammar determines the phonetic form of a sentence generated by the syntactic rules. That is, it relates a structure generated by the syntactic component to a phonetically represented signal. The semantic component determines the semantic interpretation of a sentence. That is, it relates a structure generated by the syntactic component to a certain semantic representation. Both the phonological and semantic components are therefore purely interpretive. Each utilizes information provided by the syntactic component concerning formatives, their inherent properties, and their interrelations in a given sentence. Consequently, the syntactic component of a grammar must specify, for each sentence, a *deep structure* that determines its semantic interpretation and a *surface structure* that determines its phonetic interpretation. The first of these is interpreted by the semantic component; the second, by the phonological component.¹²

It might be supposed that surface structure and deep structure will always be identical. In fact, one might briefly characterize the syntactic theories that have arisen in modern structural (taxonomic) linguistics as based on the assumption that deep and surface structures are actually the same (cf. Postal, 1964a, Chomsky, 1964). The central idea of transformational grammar is that they are, in general, distinct and that the surface structure is determined by repeated application of certain formal operations called "grammatical transformations" to objects of a more

elementary sort. If this is true (as I assume, henceforth), then the syntactic component must generate deep and surface structures, for each sentence, and must interrelate them. This idea has been clarified substantially in recent work, in ways that will be described later. In Chapter 3, I shall present a specific and, in part, new proposal as to precisely how it should be formulated. For the moment, it is sufficient to observe that although the Immediate Constituent analysis (labeled bracketing) of an actual string of formatives may be adequate as an account of surface structure, it is certainly not adequate as an account of deep structure. My concern in this book is primarily with deep structure and, in particular, with the elementary objects of which deep structure is constituted.

To clarify exposition, I shall use the following terminology, with occasional revisions as the discussion proceeds.

The *base* of the syntactic component is a system of rules that generate a highly restricted (perhaps finite) set of *basic strings*, each with an associated structural description called a *base Phrase-marker*. These base Phrase-markers are the elementary units of which deep structures are constituted. I shall assume that no ambiguity is introduced by rules of the base. This assumption seems to me correct, but has no important consequences for what follows here, though it simplifies exposition. Underlying each sentence of the language there is a sequence of base Phrase-markers, each generated by the base of the syntactic component. I shall refer to this sequence as the *basis* of the sentence that it underlies.

In addition to its base, the syntactic component of a generative grammar contains a *transformational* subcomponent. This is concerned with generating a sentence, with its surface structure, from its basis. Some familiarity with the operation and effects of transformational rules is henceforth presupposed.

Since the base generates only a restricted set of base Phrase-markers, most sentences will have a sequence of such objects as an underlying basis. Among the sentences with a single base Phrase-marker as basis, we can delimit a proper subset called "kernel sentences." These are sentences of a particularly simple

sort that involve a minimum of transformational apparatus in their generation. The notion "kernel sentence" has, I think, an important intuitive significance, but since kernel sentences play no distinctive role in generation or interpretation of sentences, I shall say nothing more about them here. One must be careful not to confuse kernel sentences with the basic strings that underlie them. The basic strings and base Phrase-markers do, it seems, play a distinctive and crucial role in language use.

Since transformations will not be considered here in detail, no careful distinction will be made, in the case of a sentence with a single element in its basis, between the basic string underlying this sentence and the sentence itself. In other words, at many points in the exposition I shall make the tacit simplifying (and contrary-to-fact) assumption that the underlying basic string is the sentence, in this case, and that the base Phrase-marker is the surface structure as well as the deep structure. I shall try to select examples in such a way as to minimize possible confusion, but the simplifying assumption should be borne in mind throughout.

§ 4. JUSTIFICATION OF GRAMMARS

Before entering directly into an investigation of the syntactic component of a generative grammar, it is important to give some thought to several methodological questions of justification and adequacy.

There is, first of all, the question of how one is to obtain information about the speaker-hearer's competence, about his knowledge of the language. Like most facts of interest and importance, this is neither presented for direct observation nor extractable from data by inductive procedures of any known sort. Clearly, the actual data of linguistic performance will provide much evidence for determining the correctness of hypotheses about underlying linguistic structure, along with introspective reports (by the native speaker, or the linguist who has learned the language). This is the position that is universally adopted in practice, although there are methodological discus-

sions that seem to imply a reluctance to use observed performance or introspective reports as evidence for some underlying reality.

In brief, it is unfortunately the case that no adequate formalizable techniques are known for obtaining reliable information concerning the facts of linguistic structure (nor is this particularly surprising). There are, in other words, very few reliable experimental or data-processing procedures for obtaining significant information concerning the linguistic intuition of the native speaker. It is important to bear in mind that when an operational procedure is proposed, it must be tested for adequacy (exactly as a theory of linguistic intuition — a grammar — must be tested for adequacy) by measuring it against the standard provided by the tacit knowledge that it attempts to specify and describe. Thus a proposed operational test for, say, segmentation into words, must meet the empirical condition of conforming, in a mass of crucial and clear cases, to the linguistic intuition of the native speaker concerning such elements. Otherwise, it is without value. The same, obviously, is true in the case of any proposed operational procedure or any proposed grammatical description. If operational procedures were available that met this test, we might be justified in relying on their results in unclear and difficult cases. This remains a hope for the future rather than a present reality, however. This is the objective situation of present-day linguistic work; allusions to presumably well-known "procedures of elicitation" or "objective methods" simply obscure the actual situation in which linguistic work must, for the present, proceed. Furthermore, there is no reason to expect that reliable operational criteria for the deeper and more important theoretical notions of linguistics (such as "grammaticalness" and "paraphrase") will ever be forthcoming.

Even though few reliable operational procedures have been developed, the theoretical (that is, grammatical) investigation of the knowledge of the native speaker can proceed perfectly well. The critical problem for grammatical theory today is not a paucity of evidence but rather the inadequacy of present theories of language to account for masses of evidence that are hardly

open to serious question. The problem for the grammarian is to construct a description and, where possible, an explanation for the enormous mass of unquestionable data concerning the linguistic intuition of the native speaker (often, himself); the problem for one concerned with operational procedures is to develop tests that give the correct results and make relevant distinctions. Neither the study of grammar nor the attempt to develop useful tests is hampered by lack of evidence with which to check results, for the present. We may hope that these efforts will converge, but they must obviously converge on the tacit knowledge of the native speaker if they are to be of any significance.

One may ask whether the necessity for present-day linguistics to give such priority to introspective evidence and to the linguistic intuition of the native speaker excludes it from the domain of science. The answer to this essentially terminological question seems to have no bearing at all on any serious issue. At most, it determines how we shall denote the kind of research that can be effectively carried out in the present state of our technique and understanding. However, this terminological question actually does relate to a different issue of some interest, namely the question whether the important feature of the successful sciences has been their search for insight or their concern for objectivity. The social and behavioral sciences provide ample evidence that objectivity can be pursued with little consequent gain in insight and understanding. On the other hand, a good case can be made for the view that the natural sciences have, by and large, sought objectivity primarily insofar as it is a tool for gaining insight (for providing phenomena that can suggest or test deeper explanatory hypotheses).

In any event, at a given stage of investigation, one whose concern is for insight and understanding (rather than for objectivity as a goal in itself) must ask whether or to what extent a wider range and more exact description of phenomena is relevant to solving the problems that he faces. In linguistics, it seems to me that sharpening of the data by more objective tests is a matter of small importance for the problems at hand. One who disagrees with this estimate of the present situation in linguistics can

justify his belief in the current importance of more objective operational tests by showing how they can lead to new and deeper understanding of linguistic structure. Perhaps the day will come when the kinds of data that we now can obtain in abundance will be insufficient to resolve deeper questions concerning the structure of language. However, many questions that can realistically and significantly be formulated today do not demand evidence of a kind that is unavailable or unattainable without significant improvements in objectivity of experimental technique.

Although there is no way to avoid the traditional assumption that the speaker-hearer's linguistic intuition is the ultimate standard that determines the accuracy of any proposed grammar, linguistic theory, or operational test, it must be emphasized, once again, that this tacit knowledge may very well not be immediately available to the user of the language. To eliminate what has seemed to some an air of paradox in this remark, let me illustrate with a few examples.

If a sentence such as "flying planes can be dangerous" is presented in an appropriately constructed context, the listener will interpret it immediately in a unique way, and will fail to detect the ambiguity. In fact, he may reject the second interpretation, when this is pointed out to him, as forced or unnatural (independently of which interpretation he originally selected under contextual pressure). Nevertheless, his intuitive knowledge of the language is clearly such that both of the interpretations (corresponding to "flying planes are dangerous" and "flying planes is dangerous") are assigned to the sentence by the grammar he has internalized in some form.

In the case just mentioned, the ambiguity may be fairly transparent. But consider such a sentence as

(5) I had a book stolen

Few hearers may be aware of the fact that their internalized grammar in fact provides at least three structural descriptions for this sentence. Nevertheless, this fact can be brought to consciousness by consideration of slight elaborations of sentence

(5), for example: (i) "I had a book stolen from my car when I stupidly left the window open," that is, "someone stole a book from my car"; (ii) "I had a book stolen from his library by a professional thief who I hired to do the job," that is, "I had someone steal a book"; (iii) "I almost had a book stolen, but they caught me leaving the library with it," that is, "I had almost succeeded in stealing a book." In bringing to consciousness the triple ambiguity of (5) in this way, we present no new information to the hearer and teach him nothing new about his language but simply arrange matters in such a way that his linguistic intuition, previously obscured, becomes evident to him.

As a final illustration, consider the sentences

(6) I persuaded John to leave

(7) I expected John to leave

The first impression of the hearer may be that these sentences receive the same structural analysis. Even fairly careful thought may fail to show him that his internalized grammar assigns very different syntactic descriptions to these sentences. In fact, so far as I have been able to discover, no English grammar has pointed out the fundamental distinction between these two constructions (in particular, my own sketches of English grammar in Chomsky, 1955, 1962*a*, failed to note this). However, it is clear that the sentences (6) and (7) are not parallel in structure. The difference can be brought out by consideration of the sentences

(8) (i) I persuaded a specialist to examine John

(ii) I persuaded John to be examined by a specialist

(9) (i) I expected a specialist to examine John

(ii) I expected John to be examined by a specialist

The sentences (9i) and (9ii) are "cognitively synonymous": one is true if and only if the other is true. But no variety of even weak paraphrase holds between (8i) and (8ii). Thus (8i) can be true or false quite independently of the truth or falsity of (8ii). Whatever difference of connotation or "topic" or emphasis one may find between (9i) and (9ii) is just the difference that exists be-

tween the active sentence "a specialist will examine John" and its passive counterpart "John will be examined by a specialist." This is not at all the case with respect to (8), however. In fact, the underlying deep structure for (6) and (8ii) must show that "John" is the Direct-Object of the Verb Phrase as well as the grammatical Subject of the embedded sentence. Furthermore, in (8ii) "John" is the logical Direct-Object of the embedded sentence, whereas in (8i) the phrase "a specialist" is the Direct-Object of the Verb Phrase and the logical Subject of the embedded sentence. In (7), (9i), and (9ii), however, the Noun Phrases "John," "a specialist," and "John," respectively, have no grammatical functions other than those that are internal to the embedded sentence; in particular, "John" is the logical Direct-Object and "a specialist" the logical Subject in the embedded sentences of (9). Thus the underlying deep structures for (8i), (8ii), (9i), and (9ii) are, respectively, the following:¹³

(10) (i) Noun Phrase — Verb — Noun Phrase — Sentence

(I — persuaded — a specialist — a specialist will examine John)

(ii) Noun Phrase — Verb — Noun Phrase — Sentence

(I — persuaded — John — a specialist will examine John)

(11) (i) Noun Phrase — Verb — Sentence

(I — expected — a specialist will examine John)

(ii) Noun Phrase — Verb — Sentence

(I — expected — a specialist will examine John)

In the case of (10ii) and (11ii), the passive transformation will apply to the embedded sentence, and in all four cases other operations will give the final surface forms of (8) and (9). The important point in the present connection is that (8i) differs from (8ii) in underlying structure, although (9i) and (9ii) are essentially the same in underlying structure. This accounts for the difference in meaning. Notice, in support of this difference in analysis, that we can have "I persuaded John that (of the fact that) Sentence," but not "I expected John that (of the fact that) Sentence."

The example (6)–(7) serves to illustrate two important points. First, it shows how unrevealing surface structure may be as to underlying deep structure. Thus (6) and (7) are the same in surface structure, but very different in the deep structure that underlies them and determines their semantic interpretations. Second, it illustrates the elusiveness of the speaker's tacit knowledge. Until such examples as (8) and (9) are adduced, it may not be in the least clear to a speaker of English that the grammar that he has internalized in fact assigns very different syntactic analyses to the superficially analogous sentences (6) and (7).

In short, we must be careful not to overlook the fact that surface similarities may hide underlying distinctions of a fundamental nature, and that it may be necessary to guide and draw out the speaker's intuition in perhaps fairly subtle ways before we can determine what is the actual character of his knowledge of his language or of anything else. Neither point is new (the former is a commonplace of traditional linguistic theory and analytic philosophy; the latter is as old as Plato's *Meno*); both are too often overlooked.

A grammar can be regarded as a theory of a language; it is *descriptively adequate* to the extent that it correctly describes the intrinsic competence of the idealized native speaker. The structural descriptions assigned to sentences by the grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be immediately aware of this) in a substantial and significant class of crucial cases.

A linguistic theory must contain a definition of "grammar," that is, a specification of the class of potential grammars. We may, correspondingly, say that a *linguistic theory is descriptively adequate* if it makes a descriptively adequate grammar available for each natural language.

Although even descriptive adequacy on a large scale is by no means easy to approach, it is crucial for the productive development of linguistic theory that much higher goals than this be pursued. To facilitate the clear formulation of deeper questions,

it is useful to consider the abstract problem of constructing an "acquisition model" for language, that is, a theory of language learning or grammar construction. Clearly, a child who has learned a language has developed an internal representation of a system of rules that determine how sentences are to be formed, used, and understood. Using the term "grammar" with a systematic ambiguity (to refer, first, to the native speaker's internally represented "theory of his language" and, second, to the linguist's account of this), we can say that the child has developed and internally represented a generative grammar, in the sense described. He has done this on the basis of observation of what we may call *primary linguistic data*. This must include examples of linguistic performance that are taken to be well-formed sentences, and may include also examples designated as non-sentences, and no doubt much other information of the sort that is required for language learning, whatever this may be (see pp. 31–32). On the basis of such data, the child constructs a grammar — that is, a theory of the language of which the well-formed sentences of the primary linguistic data constitute a small sample.¹⁴ To learn a language, then, the child must have a method for devising an appropriate grammar, given primary linguistic data. As a precondition for language learning, he must possess, first, a linguistic theory that specifies the form of the grammar of a possible human language, and, second, a strategy for selecting a grammar of the appropriate form that is compatible with the primary linguistic data. As a long-range task for general linguistics, we might set the problem of developing an account of this innate linguistic theory that provides the basis for language learning. (Note that we are again using the term "theory" — in this case "theory of language" rather than "theory of a particular language" — with a systematic ambiguity, to refer both to the child's innate predisposition to learn a language of a certain type and to the linguist's account of this.)

To the extent that a linguistic theory succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data, we can say that it meets the condition of *explanatory adequacy*. That is, to this extent, it offers an explanation for the

intuition of the native speaker on the basis of an empirical hypothesis concerning the innate predisposition of the child to develop a certain kind of theory to deal with the evidence presented to him. Any such hypothesis can be falsified (all too easily, in actual fact) by showing that it fails to provide a descriptively adequate grammar for primary linguistic data from some other language—evidently the child is not predisposed to learn one language rather than another. It is supported when it does provide an adequate explanation for some aspect of linguistic structure, an account of the way in which such knowledge might have been obtained.

Clearly, it would be utopian to expect to achieve explanatory adequacy on a large scale in the present state of linguistics. Nevertheless, considerations of explanatory adequacy are often critical for advancing linguistic theory. Gross coverage of a large mass of data can often be attained by conflicting theories; for precisely this reason it is not, in itself, an achievement of any particular theoretical interest or importance. As in any other field, the important problem in linguistics is to discover a complex of data that differentiates between conflicting conceptions of linguistic structure in that one of these conflicting theories can describe these data only by *ad hoc* means whereas the other can explain it on the basis of some empirical assumption about the form of language. Such small-scale studies of explanatory adequacy have, in fact, provided most of the evidence that has any serious bearing on the nature of linguistic structure. Thus whether we are comparing radically different theories of grammar or trying to determine the correctness of some particular aspect of one such theory, it is questions of explanatory adequacy that must, quite often, bear the burden of justification. This remark is in no way inconsistent with the fact that explanatory adequacy on a large scale is out of reach, for the present. It simply brings out the highly tentative character of any attempt to justify an empirical claim about linguistic structure.

To summarize briefly, there are two respects in which one can speak of "justifying a generative grammar." On one level (that

of descriptive adequacy), the grammar is justified to the extent that it correctly describes its object, namely the linguistic intuition—the tacit competence—of the native speaker. In this sense, the grammar is justified on *external* grounds, on grounds of correspondence to linguistic fact. On a much deeper and hence much more rarely attainable level (that of explanatory adequacy), a grammar is justified to the extent that it is a *principled* descriptively adequate system, in that the linguistic theory with which it is associated selects this grammar over others, given primary linguistic data with which all are compatible. In this sense, the grammar is justified on *internal* grounds, on grounds of its relation to a linguistic theory that constitutes an explanatory hypothesis about the form of language as such. The problem of internal justification—of explanatory adequacy—is essentially the problem of constructing a theory of language acquisition, an account of the specific innate abilities that make this achievement possible.

§ 5. FORMAL AND SUBSTANTIVE UNIVERSALS

A theory of linguistic structure that aims for explanatory adequacy incorporates an account of linguistic universals, and it attributes tacit knowledge of these universals to the child. It proposes, then, that the child approaches the data with the presumption that they are drawn from a language of a certain antecedently well-defined type, his problem being to determine which of the (humanly) possible languages is that of the community in which he is placed. Language learning would be impossible unless this were the case. The important question is: What are the initial assumptions concerning the nature of language that the child brings to language learning, and how detailed and specific is the innate schema (the general definition of "grammar") that gradually becomes more explicit and differentiated as the child learns the language? For the present we cannot come at all close to making a hypothesis about innate schemata that is rich, detailed, and specific enough to account for the fact of language acquisition. Consequently, the main

task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning.

The study of linguistic universals is the study of the properties of any generative grammar for a natural language. Particular assumptions about linguistic universals may pertain to either the syntactic, semantic, or phonological component, or to interrelations among the three components.

It is useful to classify linguistic universals as *formal* or *substantive*. A theory of substantive universals claims that items of a particular kind in any language must be drawn from a fixed class of items. For example, Jakobson's theory of distinctive features can be interpreted as making an assertion about substantive universals with respect to the phonological component of a generative grammar. It asserts that each output of this component consists of elements that are characterized in terms of some small number of fixed, universal, phonetic features (perhaps on the order of fifteen or twenty), each of which has a substantive acoustic-articulatory characterization independent of any particular language. Traditional universal grammar was also a theory of substantive universals, in this sense. It not only put forth interesting views as to the nature of universal phonetics, but also advanced the position that certain fixed syntactic categories (Noun, Verb, etc.) can be found in the syntactic representations of the sentences of any language, and that these provide the general underlying syntactic structure of each language. A theory of substantive semantic universals might hold for example, that certain designative functions must be carried out in a specified way in each language. Thus it might assert that each language will contain terms that designate persons or lexical items referring to certain specific kinds of objects, feelings, behavior, and so on.

It is also possible, however, to search for universal properties

of a more abstract sort. Consider a claim that the grammar of every language meets certain specified formal conditions. The truth of this hypothesis would not in itself imply that any particular rule must appear in all or even in any two grammars. The property of having a grammar meeting a certain abstract condition might be called a *formal* linguistic universal, if shown to be a general property of natural languages. Recent attempts to specify the abstract conditions that a generative grammar must meet have produced a variety of proposals concerning formal universals, in this sense. For example, consider the proposal that the syntactic component of a grammar must contain transformational rules (these being operations of a highly special kind) mapping semantically interpreted deep structures into phonetically interpreted surface structures, or the proposal that the phonological component of a grammar consists of a sequence of rules, a subset of which may apply cyclically to successively more dominant constituents of the surface structure (a transformational cycle, in the sense of much recent work on phonology). Such proposals make claims of a quite different sort from the claim that certain substantive phonetic elements are available for phonetic representation in all languages, or that certain specific categories must be central to the syntax of all languages, or that certain semantic features or categories provide a universal framework for semantic description. Substantive universals such as these concern the vocabulary for the description of language; formal universals involve rather the character of the rules that appear in grammars and the ways in which they can be interconnected.

On the semantic level, too, it is possible to search for what might be called formal universals, in essentially the sense just described. Consider, for example, the assumption that proper names, in any language, must designate objects meeting a condition of spatiotemporal contiguity,¹⁵ and that the same is true of other terms designating objects; or the condition that the color words of any language must subdivide the color spectrum into continuous segments; or the condition that artifacts are defined in terms of certain human goals, needs, and functions instead of solely in terms of physical qualities.¹⁶ Formal con-

straints of this sort on a system of concepts may severely limit the choice (by the child, or the linguist) of a descriptive grammar, given primary linguistic data.

The existence of deep-seated formal universals, in the sense suggested by such examples as these, implies that all languages are cut to the same pattern, but does not imply that there is any point by point correspondence between particular languages. It does not, for example, imply that there must be some reasonable procedure for translating between languages.¹⁷

In general, there is no doubt that a theory of language, regarded as a hypothesis about the innate "language-forming capacity" of humans, should concern itself with both substantive and formal universals. But whereas substantive universals have been the traditional concern of general linguistic theory, investigations of the abstract conditions that must be satisfied by any generative grammar have been undertaken only quite recently. They seem to offer extremely rich and varied possibilities for study in all aspects of grammar.

§ 6. FURTHER REMARKS ON DESCRIPTIVE AND EXPLANATORY THEORIES

Let us consider with somewhat greater care just what is involved in the construction of an "acquisition model" for language. A child who is capable of language learning must have

- (12) (i) a technique for representing input signals
- (ii) a way of representing structural information about these signals
- (iii) some initial delimitation of a class of possible hypotheses about language structure
- (iv) a method for determining what each such hypothesis implies with respect to each sentence
- (v) a method for selecting one of the (presumably, infinitely many) hypotheses that are allowed by (iii) and are compatible with the given primary linguistic data

Correspondingly, a theory of linguistic structure that aims for explanatory adequacy must contain

- (13) (i) a universal phonetic theory that defines the notion "possible sentence"
- (ii) a definition of "structural description"
- (iii) a definition of "generative grammar"
- (iv) a method for determining the structural description of a sentence, given a grammar
- (v) a way of evaluating alternative proposed grammars

Putting the same requirements in somewhat different terms, we must require of such a linguistic theory that it provide for

- (14) (i) an enumeration of the class s_1, s_2, \dots of possible sentences
- (ii) an enumeration of the class SD_1, SD_2, \dots of possible structural descriptions
- (iii) an enumeration of the class G_1, G_2, \dots of possible generative grammars
- (iv) specification of a function f such that $SD_{f(i,j)}$ is the structural description assigned to sentence s_i by grammar G_j , for arbitrary i, j ¹⁸
- (v) specification of a function m such that $m(i)$ is an integer associated with the grammar G_i as its value (with, let us say, lower value indicated by higher number)

Conditions of at least this strength are entailed by the decision to aim for explanatory adequacy.

A theory meeting these conditions would attempt to account for language learning in the following way. Consider first the nature of primary linguistic data. This consists of a finite amount of information about sentences, which, furthermore, must be rather restricted in scope, considering the time limitations that are in effect, and fairly degenerate in quality (cf. note 14). For example, certain signals might be accepted as properly formed sentences, while others are classed as nonsentences, as a result of correction of the learner's attempts on the part of the linguistic community. Furthermore, the conditions of use might be such

as to require that structural descriptions be assigned to these objects in certain ways. That the latter is a prerequisite for language acquisition seems to follow from the widely accepted (but, for the moment, quite unsupported) view that there must be a partially semantic basis for the acquisition of syntax or for the justification of hypotheses about the syntactic component of a grammar. Incidentally, it is often not realized how strong a claim this is about the innate concept-forming abilities of the child and the system of linguistic universals that these abilities imply. Thus what is maintained, presumably, is that the child has an innate theory of potential structural descriptions that is sufficiently rich and fully developed so that he is able to determine, from a real situation in which a signal occurs, which structural descriptions may be appropriate to this signal, and also that he is able to do this in part in advance of any assumption as to the linguistic structure of this signal. To say that the assumption about innate capacity is extremely strong is, of course, not to say that it is incorrect. Let us, in any event, assume tentatively that the primary linguistic data consist of signals classified as sentences and nonsentences, and a partial and tentative pairing of signals with structural descriptions.

A language-acquisition device that meets conditions (i)–(iv) is capable of utilizing such primary linguistic data as the empirical basis for language learning. This device must search through the set of possible hypotheses G_1, G_2, \dots , which are available to it by virtue of condition (iii), and must select grammars that are compatible with the primary linguistic data, represented in terms of (i) and (ii). It is possible to test compatibility by virtue of the fact that the device meets condition (iv). The device would then select one of these potential grammars by the evaluation measure guaranteed by (v).¹⁹ The selected grammar now provides the device with a method for interpreting an arbitrary sentence, by virtue of (ii) and (iv). That is to say, the device has now constructed a theory of the language of which the primary linguistic data are a sample. The theory that the device has now selected and internally represented specifies its tacit competence, its knowledge of the language. The child who acquires a language

in this way of course knows a great deal more than he has "learned." His knowledge of the language, as this is determined by his internalized grammar, goes far beyond the presented primary linguistic data and is in no sense an "inductive generalization" from these data.

This account of language learning can, obviously, be paraphrased directly as a description of how the linguist whose work is guided by a linguistic theory meeting conditions (i)–(v) would justify a grammar that he constructs for a language on the basis of given primary linguistic data.²⁰

Notice, incidentally, that care must be taken to distinguish several different ways in which primary linguistic data may be necessary for language learning. In part, such data determine to which of the possible languages (that is, the languages provided with grammars in accordance with the a priori constraint (iii)) the language learner is being exposed, and it is this function of the primary linguistic data that we are considering here. But such data may play an entirely different role as well; namely, certain kinds of data and experience may be required in order to set the language-acquisition device into operation, although they may not affect the manner of its functioning in the least. Thus it has been found that semantic reference may greatly facilitate performance in a syntax-learning experiment, even though it does not, apparently, affect the *manner* in which acquisition of syntax proceeds; that is, it plays no role in determining which hypotheses are selected by the learner (Miller and Norman, 1964). Similarly, it would not be at all surprising to find that normal language learning requires use of language in real-life situations, in some way. But this, if true, would not be sufficient to show that information regarding situational context (in particular, a pairing of signals with structural descriptions that is at least in part prior to assumptions about syntactic structure) plays any role in determining how language is acquired, once the mechanism is put to work and the task of language learning is undertaken by the child. This distinction is quite familiar outside of the domain of language acquisition. For example, Richard Held has shown in numerous experiments

that under certain circumstances reafferent stimulation (that is, stimulation resulting from voluntary activity) is a prerequisite to the development of a concept of visual space, although it may not determine the character of this concept (cf. Held and Hein, 1963; Held and Freedman, 1963, and references cited there). Or, to take one of innumerable examples from studies of animal learning, it has been observed (Lemmon and Patterson, 1964) that depth perception in lambs is considerably facilitated by mother-neonate contact, although again there is no reason to suppose that the nature of the lamb's "theory of visual space" depends on this contact.

In studying the actual character of learning, linguistic or otherwise, it is of course necessary to distinguish carefully between these two functions of external data — the function of initiating or facilitating the operation of innate mechanisms and the function of determining in part the direction that learning will take.²¹

Returning now to the main theme, we shall call a theory of linguistic structure that meets conditions (i)–(v) an *explanatory theory*, and a theory that meets conditions (i)–(iv) a *descriptive theory*. In fact, a linguistic theory that is concerned only with descriptive adequacy will limit its attention to topics (i)–(iv). Such a theory must, in other words, make available a class of generative grammars containing, for each language, a descriptively adequate grammar of this language — a grammar that (by means of (iv)) assigns structural descriptions to sentences in accordance with the linguistic competence of the native speaker. A theory of language is empirically significant only to the extent that it meets conditions (i)–(iv). The further question of explanatory adequacy arises only in connection with a theory that also meets condition (v) (but see p. 36). In other words, it arises only to the extent that the theory provides a principled basis for selecting a descriptively adequate grammar on the basis of primary linguistic data by the use of a well-defined evaluation measure.

This account is misleading in one important respect. It suggests that to raise a descriptively adequate theory to the level

of explanatory adequacy one needs only to define an appropriate evaluation measure. This is incorrect, however. A theory may be descriptively adequate, in the sense just defined, and yet provide such a wide range of potential grammars that there is no possibility of discovering a formal property distinguishing the descriptively adequate grammars, in general, from among the mass of grammars compatible with whatever data are available. In fact, the real problem is almost always to restrict the range of possible hypotheses by adding additional structure to the notion "generative grammar." For the construction of a reasonable acquisition model, it is necessary to reduce the class of attainable²² grammars compatible with given primary linguistic data to the point where selection among them can be made by a formal evaluation measure. This requires a precise and narrow delimitation of the notion "generative grammar" — a restrictive and rich hypothesis concerning the universal properties that determine the form of language, in the traditional sense of this term.

The same point can be put in a somewhat different way. Given a variety of descriptively adequate grammars for natural languages, we are interested in determining to what extent they are unique and to what extent there are deep underlying similarities among them that are attributable to the form of language as such. Real progress in linguistics consists in the discovery that certain features of given languages can be reduced to universal properties of language, and explained in terms of these deeper aspects of linguistic form. Thus the major endeavor of the linguist must be to enrich the theory of linguistic form by formulating more specific constraints and conditions on the notion "generative grammar." Where this can be done, particular grammars can be simplified by eliminating from them descriptive statements that are attributable to the general theory of grammar (cf. § 5). For example, if we conclude that the transformational cycle²³ is a universal feature of the phonological component, it is unnecessary, in the grammar of English, to describe the manner of functioning of those phonological rules that involve syntactic structure. This description will now have been ab-

stracted from the grammar of English and stated as a formal linguistic universal, as part of the theory of generative grammar. Obviously, this conclusion, if justified, would represent an important advance in the theory of language, since it would then have been shown that what appears to be a peculiarity of English is actually explicable in terms of a general and deep empirical assumption about the nature of language, an assumption that can be refuted, if false, by study of descriptively adequate grammars of other languages.

In short, the most serious problem that arises in the attempt to achieve explanatory adequacy is that of characterizing the notion "generative grammar" in a sufficiently rich, detailed, and highly structured way. A theory of grammar may be descriptively adequate and yet leave unexpressed major features that are defining properties of natural language and that distinguish natural languages from arbitrary symbolic systems. It is for just this reason that the attempt to achieve explanatory adequacy — the attempt to discover linguistic universals — is so crucial at every stage of understanding of linguistic structure, despite the fact that even descriptive adequacy on a broad scale may be an unrealized goal. It is not necessary to achieve descriptive adequacy before raising questions of explanatory adequacy. On the contrary, the crucial questions, the questions that have the greatest bearing on our concept of language and on descriptive practice as well, are almost always those involving explanatory adequacy with respect to particular aspects of language structure.

To acquire language, a child must devise a hypothesis compatible with presented data — he must select from the store of potential grammars a specific one that is appropriate to the data available to him. It is logically possible that the data might be sufficiently rich and the class of potential grammars sufficiently limited so that no more than a single permitted grammar will be compatible with the available data at the moment of successful language acquisition, in our idealized "instantaneous" model (cf. notes 19 and 22). In this case, no evaluation procedure will be necessary as a part of linguistic theory — that is, as an innate property of an organism or a device capable of language acquisi-

tion. It is rather difficult to imagine how in detail this logical possibility might be realized, and all concrete attempts to formulate an empirically adequate linguistic theory certainly leave ample room for mutually inconsistent grammars, all compatible with primary data of any conceivable sort. All such theories therefore require supplementation by an evaluation measure if language acquisition is to be accounted for and selection of specific grammars is to be justified; and I shall continue to assume tentatively, as heretofore, that this is an empirical fact about the innate human *faculté de langage* and consequently about general linguistic theory as well.

§ 7. ON EVALUATION PROCEDURES

The status of an evaluation procedure for grammars (see condition (v) of (12)–(14)) has often been misconstrued. It must first of all be kept clearly in mind that such a measure is not given a priori, in some manner. Rather, any proposal concerning such a measure is an empirical hypothesis about the nature of language. This is evident from the preceding discussion. Suppose that we have a descriptive theory, meeting conditions (i)–(iv) of (12)–(14) in some fixed way. Given primarily linguistic data *D*, different choices of an evaluation measure will assign quite different ranks to alternative hypotheses (alternative grammars) as to the language of which *D* is a sample, and will therefore lead to entirely different predictions as to how a person who learns a language on the basis of *D* will interpret new sentences not in *D*. Consequently, choice of an evaluation measure is an empirical matter, and particular proposals are correct or incorrect.

Perhaps confusion about this matter can be traced to the use of the term "simplicity measure" for particular proposed evaluation measures, it being assumed that "simplicity" is a general notion somehow understood in advance outside of linguistic theory. This is a misconception, however. In the context of this discussion, "simplicity" (that is, the evaluation measure *m* of (v)) is a notion to be defined within linguistic theory along with "grammar," "phoneme," etc. Choice of a simplicity measure is

rather like determination of the value of a physical constant. We are given, in part, an empirical pairing of certain kinds of primary linguistic data with certain grammars that are in fact constructed by people presented with such data. A proposed simplicity measure constitutes part of the attempt to determine precisely the nature of this association. If a particular formulation of (i)–(iv) is assumed, and if pairs (D_1, G_1) , (D_2, G_2) , ... of primary linguistic data and descriptively adequate grammars are given, the problem of defining "simplicity" is just the problem of discovering how G_i is determined by D_i , for each i . Suppose, in other words, that we regard an acquisition model for language as an input-output device that determines a particular generative grammar as "output," given certain primary linguistic data as input. A proposed simplicity measure, taken together with a specification of (i)–(iv), constitutes a hypothesis concerning the nature of such a device. Choice of a simplicity measure is therefore an empirical matter with empirical consequences.

All of this has been said before. I repeat it at such length because it has been so grossly misunderstood.

It is also apparent that evaluation measures of the kinds that have been discussed in the literature on generative grammar cannot be used to compare different theories of grammar; comparison of a grammar from one class of proposed grammars with a grammar from another class, *by such a measure*, is utterly without sense. Rather, an evaluation measure of this kind is an essential part of a particular theory of grammar that aims at explanatory adequacy. It is true that there is a sense in which alternative theories of language (or alternative theories in other domains) can be compared as to simplicity and elegance. What we have been discussing here, however, is not this general question but rather the problem of comparing two theories of a language — two grammars of this language — in terms of a particular general linguistic theory. This is, then, a matter of formulating an explanatory theory of language; it is not to be confused with the problem of choosing among competing theories of language. Choice among competing theories of language is of course a fundamental question and should also be

settled, insofar as possible, on empirical grounds of descriptive and explanatory adequacy. But it is not the question involved in the use of an evaluation measure in the attempt to achieve explanatory adequacy.

As a concrete illustration, consider the question of whether the rules of a grammar should be unordered (let us call this the linguistic theory T_U) or ordered in some specific way (the theory T_O). A priori, there is no way to decide which of the two is correct. There is no known absolute sense of "simplicity" or "elegance," developed within linguistic theory or general epistemology, in accordance with which T_U and T_O can be compared. It is quite meaningless, therefore, to maintain that in some absolute sense T_U is "simpler" than T_O or conversely. One can easily invent a general concept of "simplicity" that will prefer T_U to T_O , or T_O to T_U ; in neither case will this concept have any known justification. Certain measures of evaluation have been proposed and in part empirically justified within linguistics — for example, minimization of feature specification (as discussed in Halle, 1959a, 1961, 1962a, 1964) or the measure based on abbreviatory notations (discussed on pp. 42f.). These measures do not apply, because they are internal to a specific linguistic theory and their empirical justification relies essentially on this fact. To choose between T_U and T_O , we must proceed in an entirely different way. We must ask whether T_U or T_O provides descriptively adequate grammars for natural languages, or leads to explanatory adequacy. This is a perfectly meaningful empirical question if the theories in question are stated with sufficient care. For example, if T_U^s is the familiar theory of phrase structure grammar and T_O^s is the same theory, with the further condition that the rules are linearly ordered and apply cyclically, with at least one rule $A \rightarrow X$ being obligatory for each category A , so as to guarantee that each cycle is nonvacuous, then it can be shown that T_U^s and T_O^s are incomparable in descriptive power (in "strong generative capacity" — see § 9; see Chomsky, 1955, Chapters 6 and 7, and Chomsky, 1956, for some discussion of such systems). Consequently, we might ask whether natural languages in fact fall under T_U^s or T_O^s , these being non-

equivalent and empirically distinguishable theories. Or, supposing T_V^P and T_O^P to be theories of the phonological component (where T_V^P holds phonological rules to be unordered and T_O^P holds them to be partially ordered), it is easy to invent hypothetical "languages" for which significant generalizations are expressible in terms of T_O^P but not T_V^P , or conversely. We can therefore try to determine whether there are significant generalizations that are expressible in terms of one but not the other theory in the case of empirically given languages. In principle, either result is possible; it is an entirely factual question, having to do with the properties of natural languages. We shall see later that T_O^S is rather well motivated as a theory of the base, and strong arguments have been offered to show that T_O^P is correct and T_V^P is wrong, as a theory of phonological processes (cf. Chomsky, 1951, 1964; Halle, 1959a, 1959b, 1962a, 1964). In both cases, the argument turns on the factual question of expressibility of linguistically significant generalizations in terms of one or the other theory, not on any presumed absolute sense of "simplicity" that might rank T_V and T_O relative to one another. Failure to appreciate this fact has led to a great deal of vacuous and pointless discussion.

Confusion about these questions may also have been engendered by the fact that there are several different senses in which one can talk of "justifying" a grammar, as noted on pp. 26-27. To repeat the major point: on the one hand, the grammar can be justified on external grounds of descriptive adequacy—we may ask whether it states the facts about the language correctly, whether it predicts correctly how the idealized native speaker would understand arbitrary sentences and gives a correct account of the basis for this achievement; on the other hand, a grammar can be justified on internal grounds if, given an explanatory linguistic theory, it can be shown that this grammar is the highest-valued grammar permitted by the theory and compatible with given primary linguistic data. In the latter case, a principled basis is presented for the construction of this grammar, and it is therefore justified on much deeper empirical grounds. Both kinds of justification are of course necessary; it is im-

portant, however, not to confuse them. In the case of a linguistic theory that is merely descriptive, only one kind of justification can be given—namely, we can show that it permits grammars that meet the external condition of descriptive adequacy.²⁴ It is only when all of the conditions (i)-(v) of (12)-(14) are met that the deeper question of internal justification can be raised.

It is also apparent that the discussion as to whether an evaluation measure is a "necessary" part of linguistic theory is quite without substance (see, however, pp. 36-37). If the linguist is content to formulate descriptions one way or another with little concern for justification, and if he does not intend to proceed from the study of facts about particular languages to an investigation of the characteristic properties of natural language as such, then construction of an evaluation procedure and the associated concerns that relate to explanatory adequacy need not concern him. In this case, since interest in justification has been abandoned, neither evidence nor argument (beyond minimal requirements of consistency) has any bearing on what the linguist presents as a linguistic description. On the other hand, if he wishes to achieve descriptive adequacy in his account of language structure, he must concern himself with the problem of developing an explanatory theory of the form of grammar, since this provides one of the main tools for arriving at a descriptively adequate grammar in any particular case. In other words, choice of a grammar for a particular language L will always be much underdetermined by the data drawn from L alone. Moreover, other relevant data (namely, successful grammars for other languages or successful fragments for other subparts of L) will be available to the linguist only if he possesses an explanatory theory. Such a theory limits the choice of grammar by the dual method of imposing formal conditions on grammar and providing an evaluation procedure to be applied for the language L with which he is now concerned. Both the formal conditions and the evaluation procedure can be empirically justified by their success in other cases. Hence, any far-reaching concern for descriptive adequacy must lead to an attempt to develop an explanatory theory that fulfills these dual functions, and concern with ex-

planatory adequacy surely requires an investigation of evaluation procedures.

The major problem in constructing an evaluation measure for grammars is that of determining which generalizations about a language are significant ones; an evaluation measure must be selected in such a way as to favor these. We have a generalization when a set of rules about distinct items can be replaced by a single rule (or, more generally, partially identical rules) about the whole set, or when it can be shown that a "natural class" of items undergoes a certain process or set of similar processes. Thus, choice of an evaluation measure constitutes a decision as to what are "similar processes" and "natural classes" — in short, what are significant generalizations. The problem is to devise a procedure that will assign a numerical measure of valuation to a grammar in terms of the degree of linguistically significant generalization that this grammar achieves. The obvious numerical measure to be applied to a grammar is length, in terms of number of symbols. But if this is to be a meaningful measure, it is necessary to devise notations and to restrict the form of rules in such a way that significant considerations of complexity and generality are converted into considerations of length, so that real generalizations shorten the grammar and spurious ones do not. Thus it is the notational conventions used in presenting a grammar that define "significant generalization," if the evaluation measure is taken as length.

This is, in fact, the rationale behind the conventions for use of parentheses, brackets, etc., that have been adopted in explicit (that is, generative) grammars. For a detailed discussion of these, see Chomsky (1951, 1955), Postal (1962a), and Matthews (1964). To take just one example, consider the analysis of the English Verbal Auxiliary. The facts are that such a phrase must contain Tense (which is, furthermore, *Past* or *Present*), and then may or may not contain a Modal and either the *Perfect* or *Progressive* Aspect (or both), where the elements must appear in the order just given. Using familiar notational conventions, we can state this rule in the following form:

(15) Aux \rightarrow Tense (Modal)(*Perfect*)(*Progressive*)

(omitting details that are not relevant here). Rule (15) is an abbreviation for eight rules that analyze the element Aux into its eight possible forms. Stated in full, these eight rules would involve twenty symbols, whereas rule (15) involves four (not counting Aux, in both cases). The parenthesis notation, in this case, has the following meaning. It asserts that the difference between four and twenty symbols is a measure of the degree of linguistically significant generalization achieved in a language that has the forms given in list (16), for the Auxiliary Phrase, as compared with a language that has, for example, the forms given in list (17) as the representatives of this category:

(16) Tense, Tense \wedge Modal, Tense \wedge *Perfect*, Tense \wedge *Progressive*,
Tense \wedge Modal \wedge *Perfect*, Tense \wedge Modal \wedge *Progressive*, Tense
 \wedge *Perfect* \wedge *Progressive*, Tense \wedge Modal \wedge *Perfect* \wedge *Progressive*

(17) Tense \wedge Modal \wedge *Perfect* \wedge *Progressive*, Modal \wedge *Perfect* \wedge *Pro-*
gressive \wedge Tense, *Perfect* \wedge *Progressive* \wedge Tense \wedge Modal, *Pro-*
gressive \wedge Tense \wedge Modal \wedge *Perfect*, Tense \wedge *Perfect*, Modal \wedge
Progressive

In the case of both list (16) and list (17), twenty symbols are involved. List (16) abbreviates to rule (15) by the notational convention; list (17) cannot be abbreviated by this convention. Hence, adoption of the familiar notational conventions involving the use of parentheses amounts to a claim that there is a linguistically significant generalization underlying the set of forms in list (16) but not the set of forms in list (17). It amounts to the empirical hypothesis that regularities of the type exemplified in (16) are those found in natural languages, and are of the type that children learning a language will expect; whereas cyclic regularities of the type exemplified in (17), though perfectly genuine, abstractly, are not characteristic of natural language, are not of the type for which children will intuitively search in language materials, and are much more difficult for the language-learner to construct on the basis of scattered data

or to use. What is claimed, then, is that when given scattered examples from (16), the language learner will construct the rule (15) generating the full set with their semantic interpretations, whereas when given scattered examples that could be subsumed under a cyclic rule, he will not incorporate this "generalization" in his grammar—he will not, for example, conclude from the existence of "yesterday John arrived" and "John arrived yesterday" that there is a third form "arrived yesterday John," or from the existence of "is John here" and "here is John" that there is a third form "John here is," etc. One might easily propose a different notational convention that would abbreviate list (17) to a shorter rule than list (16), thus making a different empirical assumption about what constitutes a linguistically significant generalization. There is no a priori reason for preferring the usual convention; it simply embodies a factual claim about the structure of natural language and the predisposition of the child to search for certain types of regularity in natural language.

The illustrative examples of the preceding paragraph must be regarded with some caution. It is the full set of notational conventions that constitute an evaluation procedure, in the manner outlined earlier. The factual content of an explanatory theory lies in its claim that the most highly valued grammar of the permitted form will be selected, on the basis of given data. Hence, descriptions of particular subsystems of the grammar must be evaluated in terms of their effect on the entire system of rules. The extent to which particular parts of the grammar can be selected independently of others is an empirical matter about which very little is known, at present. Although alternatives can be clearly formulated, deeper studies of particular languages than are presently available are needed to settle the questions that immediately arise when these extremely important issues are raised. To my knowledge, the only attempt to evaluate a fairly full and complex subsystem of a grammar is in Chomsky (1951), but even here all that is shown is that the value of the system is a "local maximum" in the sense that interchange of adjacent rules decreases value. The effect of modifications on a larger

scale is not investigated. Certain aspects of the general question, relating to lexical and phonological structure, are discussed in Halle and Chomsky (forthcoming).

One special case of this general approach to evaluation that has been worked out in a particularly convincing way is the condition of minimization of distinctive feature specifications in the phonological component of the grammar. A very plausible argument can be given to the effect that this convention defines the notions of "natural class" and "significant generalization" that have been relied on implicitly in descriptive and comparative-historical phonological investigations, and that determine the intuitively given distinction between "phonologically possible" and "phonologically impossible" nonsense forms. For discussion, see Halle (1959*a*, 1959*b*, 1961, 1962*a*, 1964), Halle and Chomsky (forthcoming). It is important to observe that the effectiveness of this particular evaluation measure is completely dependent on a strong assumption about the form of grammar, namely, the assumption that only feature notation is permitted. If phonemic notation is allowed in addition to feature notation, the measure gives absurd consequences, as Halle shows.

It is clear, then, that choice of notations and other conventions is not an arbitrary or "merely technical" matter, if length is to be taken as the measure of valuation for a grammar. It is, rather, a matter that has immediate and perhaps quite drastic empirical consequences. When particular notational devices are incorporated into a linguistic theory of the sort we are discussing, a certain empirical claim is made, implicitly, concerning natural language. It is implied that a person learning a language will attempt to formulate generalizations that can easily be expressed (that is, with few symbols) in terms of the notations available in this theory, and that he will select grammars containing these generalizations over other grammars that are also compatible with the given data but that contain different sorts of generalization, different concepts of "natural class," and so on. These may be very strong claims, and need by no means be true on any a priori grounds.

To avoid any possible lingering confusion on this matter,

let me repeat once more that this discussion of language learning in terms of formulation of rules, hypotheses, etc., does not refer to conscious formulation and expression of these but rather to the process of arriving at an internal representation of a generative system, which can be appropriately described in these terms.

In brief, it is clear that no present-day theory of language can hope to attain explanatory adequacy beyond very restricted domains. In other words, we are very far from being able to present a system of formal and substantive linguistic universals that will be sufficiently rich and detailed to account for the facts of language learning. To advance linguistic theory in the direction of explanatory adequacy, we can attempt to refine the evaluation measure for grammars or to tighten the formal constraints on grammars so that it becomes more difficult to find a highly valued hypothesis compatible with primary linguistic data. There can be no doubt that present theories of grammar require modification in both of these ways, the latter, in general, being the more promising. Thus the most crucial problem for linguistic theory seems to be to abstract statements and generalizations from particular descriptively adequate grammars and, wherever possible, to attribute them to the general theory of linguistic structure, thus enriching this theory and imposing more structure on the schema for grammatical description. Whenever this is done, an assertion about a particular language is replaced by a corresponding assertion, from which the first follows, about language in general. If this formulation of a deeper hypothesis is incorrect, this fact should become evident when its effect on the description of other aspects of the language or the description of other languages is ascertained. In short, I am making the obvious comment that, wherever possible, general assumptions about the nature of language should be formulated from which particular features of the grammars of individual languages can be deduced. In this way, linguistic theory may move toward explanatory adequacy and contribute to the study of human mental processes and intellectual capacity—more specifically, to the determination of the abilities that make

language learning possible under the empirically given limitations of time and data.

§ 8. LINGUISTIC THEORY AND LANGUAGE LEARNING

In the preceding discussion, certain problems of linguistic theory have been formulated as questions about the construction of a hypothetical language-acquisition device. This seems a useful and suggestive framework within which to pose and consider these problems. We may think of the theorist as given an empirical pairing of collections of primary linguistic data associated with grammars that are constructed by the device on the basis of such data. Much information can be obtained about both the primary data that constitute the input and the grammar that is the "output" of such a device, and the theorist has the problem of determining the intrinsic properties of a device capable of mediating this input-output relation.

It may be of some interest to set this discussion in a somewhat more general and traditional framework. Historically, we can distinguish two general lines of approach to the problem of acquisition of knowledge, of which the problem of acquisition of language is a special and particularly informative case. The empiricist approach has assumed that the structure of the acquisition device is limited to certain elementary "peripheral processing mechanisms" — for example, in recent versions, an innate "quality space" with an innate "distance" defined on it (Quine, 1960, pp. 83f.),²⁵ a set of primitive unconditioned reflexes (Hull, 1943), or, in the case of language, the set of all "aurally distinguishable components" of the full "auditory impression" (Bloch, 1950). Beyond this, it assumes that the device has certain analytical data-processing mechanisms or inductive principles of a very elementary sort, for example, certain principles of association, weak principles of "generalization" involving gradients along the dimensions of the given quality space, or, in our case, taxonomic principles of segmentation and classification such as those that have been developed with some care in modern linguistics, in accordance with the Saussurian emphasis

on the fundamental character of such principles. It is then assumed that a preliminary analysis of experience is provided by the peripheral processing mechanisms, and that one's concepts and knowledge, beyond this, are acquired by application of the available inductive principles to this initially analyzed experience.²⁶ Such views can be formulated clearly in one way or another as empirical hypotheses about the nature of mind.

A rather different approach to the problem of acquisition of knowledge has been characteristic of rationalist speculation about mental processes. The rationalist approach holds that beyond the peripheral processing mechanisms,²⁷ there are innate ideas and principles of various kinds that determine the form of the acquired knowledge in what may be a rather restricted and highly organized way. A condition for innate mechanisms to become activated is that appropriate stimulation be presented. Thus for Descartes (1647), the innate ideas are those arising from the faculty of thinking rather than from external objects:

... nothing reaches our mind from external objects through the organs of sense beyond certain corporeal movements . . . but even these movements, and the figures which arise from them, are not conceived by us in the shape they assume in the organs of sense. . . . Hence it follows that the ideas of the movements and figures are themselves innate in us. So much the more must the ideas of pain, colour, sound and the like be innate, that our mind may, on occasion of certain corporeal movements, envisage these ideas, for they have no likeness to the corporeal movements . . . [p. 443].

Similarly, such notions as that things equal to the same thing are equal to each other are innate, since they cannot arise as necessary principles from "particular movements." In general,

sight . . . presents nothing beyond pictures, and hearing nothing beyond voices or sounds, so that all these things that we think of, beyond these voices or pictures, as being symbolized by them, are presented to us by means of ideas which come from no other source than our faculty of thinking, and are accordingly together with that faculty innate in us, that is, always existing in us potentially; for existence in any faculty is not actual but merely potential existence, since the very word "faculty" designates nothing more or less than a potentiality. . . . [Thus

ideas are innate in the sense that] in some families generosity is innate, in others certain diseases like gout or gravel, not that on this account the babes of these families suffer from these diseases in their mother's womb, but because they are born with a certain disposition or propensity for contracting them . . . [p. 442].

Still earlier, Lord Herbert (1624) maintains that innate ideas and principles "remain latent when their corresponding objects are not present, and even disappear and give no sign of their existence"; they "must be deemed not so much the outcome of experience as principles without which we should have no experience at all . . . [p. 132]." Without these principles, "we could have no experience at all nor be capable of observations"; "we should never come to distinguish between things, or to grasp any general nature . . . [p. 105]." These notions are extensively developed throughout seventeenth-century rationalist philosophy. To mention just one example, Cudworth (1731) gives an extensive argument in support of his view that "there are many ideas of the mind, which though the cogitations of them be often occasionally invited from the motion or appulse of sensible objects without made upon our bodies; yet notwithstanding the ideas themselves could not possibly be stamped or impressed upon the soul from them, because sense takes no cognizance at all of any such things in those corporeal objects, and therefore they must needs arise from the innate vigour and activity of the mind itself . . . [Book IV]." Even in Locke one finds essentially the same conception, as was pointed out by Leibniz and many commentators since.

In the Port-Royal *Logic* (Arnauld, 1662), the same point of view is expressed in the following way:

It is false, therefore, that all our ideas come through sense. On the contrary, it may be affirmed that no idea which we have in our minds has taken its rise from sense, except on occasion of those movements which are made in the brain through sense, the impulse from sense giving occasion to the mind to form different ideas which it would not have formed without it, though these ideas have very rarely any resemblance to what takes place in the sense and in the brain; and there are at least a very great number of ideas which, having no connection with any

bodily image, cannot, without manifest absurdity, be referred to sense . . . [Chapter 1].

In the same vein, Leibniz refuses to accept a sharp distinction between innate and learned:

I agree that we learn ideas and innate truths either in considering their source or in verifying them through experience. . . . And I cannot admit this proposition: *all that one learns is not innate*. The truths of numbers are in us, yet nonetheless one learns them,²⁸ either by drawing them from their source when we learn them through demonstrative proof (which shows that they are innate), or by testing them in examples, as do ordinary arithmeticians . . . [*New Essays*, p. 75]. [Thus] all arithmetic and all geometry are in us virtually, so that we can find them there if we consider attentively and set in order what we already have in the mind . . . [p. 78]. [In general,] we have an infinite amount of knowledge of which we are not always conscious, not even when we need it [p. 77]. The senses, although necessary for all our actual knowledge, are not sufficient to give it all to us, since the senses never give us anything but examples, i.e., particular or individual truths. Now all the examples which confirm a general truth, whatever their number, do not suffice to establish the universal necessity of that same truth . . . [pp. 42-43]. Necessary truths . . . must have principles whose proof does not depend on examples, nor consequently upon the testimony of the senses, although without the senses it would never have occurred to us to think of them. . . . It is true that we must not imagine that these eternal laws of the reason can be read in the soul as in an open book . . . but it is sufficient that they can be discovered in us by dint of attention, for which the senses furnish occasions, and successful experience serves to confirm reason . . . [p. 44]. [There are innate general principles that] enter into our thoughts, of which they form the soul and the connection. They are as necessary thereto as the muscles and sinews are for walking, although we do not at all think of them. The mind leans upon these principles every moment, but it does not come so easily to distinguish them and to represent them distinctly and separately, because that demands great attention to its acts. . . . Thus it is that one possesses many things without knowing it . . . [p. 74].

(as, for example, the Chinese possess articulate sounds, and therefore the basis for alphabetic writing, although they have not invented this).

Notice, incidentally, that throughout these classical discussions of the interplay between sense and mind in the formation of ideas, no sharp distinction is made between perception and acquisition, although there would be no inconsistency in the assumption that latent innate mental structures, once "activated," are then available for interpretation of the data of sense in a way in which they were not previously.

Applying this rationalist view to the special case of language learning, Humboldt (1836) concludes that one cannot really teach language but can only present the conditions under which it will develop spontaneously in the mind in its own way. Thus the *form of a language*, the schema for its grammar, is to a large extent given, though it will not be available for use without appropriate experience to set the language-forming processes into operation. Like Leibniz, he reiterates the Platonistic view that, for the individual, learning is largely a matter of *Wiederzeugung*, that is, of drawing out what is innate in the mind.²⁹

This view contrasts sharply with the empiricist notion (the prevailing modern view) that language is essentially an adventitious construct, taught by "conditioning" (as would be maintained, for example, by Skinner or Quine) or by drill and explicit explanation (as was claimed by Wittgenstein), or built up by elementary "data-processing" procedures (as modern linguistics typically maintains), but, in any event, relatively independent in its structure of any innate mental faculties.

In short, empiricist speculation has characteristically assumed that only the procedures and mechanisms for the acquisition of knowledge constitute an innate property of the mind. Thus for Hume, the method of "experimental reasoning" is a basic instinct in animals and humans, on a par with the instinct "which teaches a bird, with such exactness, the art of incubation, and the whole economy and order of its nursery"—it is derived "from the original hand of nature" (Hume, 1748, § IX). The form of knowledge, however, is otherwise quite free. On the other hand, rationalist speculation has assumed that the general form of a system of knowledge is fixed in advance as a disposition of the mind, and the function of experience is to cause this general

schematic structure to be realized and more fully differentiated. To follow Leibniz's enlightening analogy, we may make

... the comparison of a block of marble which has veins, rather than a block of marble wholly even, or of blank tablets, i.e., of what is called among philosophers a *tabula rasa*. For if the soul resembled these blank tablets, truths would be in us as the figure of Hercules is in the marble, when the marble is wholly indifferent to the reception of this figure or some other. But if there were veins in the block which should indicate the figure of Hercules rather than other figures, this block would be more determined thereto, and Hercules would be in it as in some sense innate, although it would be needful to labor to discover these veins, to clear them by polishing, and by cutting away what prevents them from appearing. Thus it is that ideas and truths are for us innate, as inclinations, dispositions, habits, or natural potentialities, and not as actions; although these potentialities are always accompanied by some actions, often insensible, which correspond to them [Leibniz, *New Essays*, pp. 45-46].

It is not, of course, necessary to assume that empiricist and rationalist views can always be sharply distinguished and that these currents cannot cross. Nevertheless, it is historically accurate as well as heuristically valuable to distinguish these two very different approaches to the problem of acquisition of knowledge. Particular empiricist and rationalist views can be made quite precise and can then be presented as explicit hypotheses about acquisition of knowledge, in particular, about the innate structure of a language-acquisition device. In fact, it would not be inaccurate to describe the taxonomic, data-processing approach of modern linguistics as an empiricist view that contrasts with the essentially rationalist alternative proposed in recent theories of transformational grammar. Taxonomic linguistics is empiricist in its assumption that general linguistic theory consists only of a body of procedures for determining the grammar of a language from a corpus of data, the form of language being unspecified except insofar as restrictions on possible grammars are determined by this set of procedures. If we interpret taxonomic linguistics as making an empirical claim,⁸⁰

this claim must be that the grammars that result from application of the postulated procedures to a sufficiently rich selection of data will be descriptively adequate—in other words, that the set of procedures can be regarded as constituting a hypothesis about the innate language-acquisition system. In contrast, the discussion of language acquisition in preceding sections was rationalistic in its assumption that various formal and substantive universals are intrinsic properties of the language-acquisition system, these providing a schema that is applied to data and that determines in a highly restricted way the general form and, in part, even the substantive features of the grammar that may emerge upon presentation of appropriate data. A general linguistic theory of the sort roughly described earlier, and elaborated in more detail in the following chapters and in other studies of transformational grammar, must therefore be regarded as a specific hypothesis, of an essentially rationalist cast, as to the nature of mental structures and processes. See Chomsky (1959*b*, 1962*b*, 1964) and Katz (forthcoming) for some further discussion of this point.

When such contrasting views are clearly formulated, we may ask, as an empirical question, which (if either) is correct. There is no a priori way to settle this issue. Where empiricist and rationalist views have been presented with sufficient care so that the question of correctness can be seriously raised, it cannot, for example, be maintained that in any clear sense one is "simpler" than the other in terms of its potential physical realization,⁸¹ and even if this could be shown, one way or the other, it would have no bearing on what is completely a factual issue. This factual question can be approached in several ways. In particular, restricting ourselves now to the question of language acquisition, we must bear in mind that any concrete empiricist proposal does impose certain conditions on the form of the grammars that can result from application of its inductive principles to primary data. We may therefore ask whether the grammars that these principles can provide, in principle, are at all close to those which we in fact discover when we investigate

real languages. The same question can be asked about a concrete rationalist proposal. This has, in the past, proved to be a useful way to subject such hypotheses to one sort of empirical test.

If the answer to this question of adequacy-in-principle is positive, in either case, we can then turn to the question of feasibility: can the inductive procedures (in the empiricist case) or the mechanisms of elaboration and realization of innate schemata (in the rationalist case) succeed in producing grammars within the given constraints of time and access, and within the range of observed uniformity of output? In fact, the second question has rarely been raised in any serious way in connection with empiricist views (but cf. Miller, Galanter, and Pribram, 1960, pp. 145-148, and Miller and Chomsky, 1963, p. 430, for some comments), since study of the first question has been sufficient to rule out whatever explicit proposals of an essentially empiricist character have emerged in modern discussions of language acquisition. The only proposals that are explicit enough to support serious study are those that have been developed within taxonomic linguistics. It seems to have been demonstrated beyond any reasonable doubt that, quite apart from any question of feasibility, methods of the sort that have been studied in taxonomic linguistics are intrinsically incapable of yielding the systems of grammatical knowledge that must be attributed to the speaker of a language (cf. Chomsky, 1956, 1957, 1964; Postal, 1962*b*, 1964*a*, 1964*c*; Katz and Postal, 1964, § 5.5, and many other publications for discussion of these questions that seems unanswerable and is, for the moment, not challenged). In general, then, it seems to me correct to say that empiricist theories about language acquisition are refutable wherever they are clear, and that further empiricist speculations have been quite empty and uninformative. On the other hand, the rationalist approach exemplified by recent work in the theory of transformational grammar seems to have proved fairly productive, to be fully in accord with what is known about language, and to offer at least some hope of providing a hypothesis about the intrinsic structure of a language-acquisition system that will meet the condition of adequacy-in-principle and do so in a sufficiently

narrow and interesting way so that the question of feasibility can, for the first time, be seriously raised.

One might seek other ways of testing particular hypotheses about a language-acquisition device. A theory that attributes possession of certain linguistic universals to a language-acquisition system, as a property to be realized under appropriate external conditions, implies that only certain kinds of symbolic systems can be acquired and used as languages by this device. Others should be beyond its language-acquisition capacity. Systems can certainly be invented that fail the conditions, formal and substantive, that have been proposed as tentative linguistic universals in, for example, Jakobsonian distinctive-feature theory or the theory of transformational grammar. In principle, one might try to determine whether invented systems that fail these conditions do pose inordinately difficult problems for language learning, and do fall beyond the domain for which the language-acquisition system is designed. As a concrete example, consider the fact that, according to the theory of transformational grammar, only certain kinds of formal operations on strings can appear in grammars — operations that, furthermore, have no a priori justification. For example, the permitted operations cannot be shown in any sense to be the most "simple" or "elementary" ones that might be invented. In fact, what might in general be considered "elementary operations" on strings do not qualify as grammatical transformations at all, while many of the operations that do qualify are far from elementary, in any general sense. Specifically, grammatical transformations are necessarily "structure-dependent" in that they manipulate substrings only in terms of their assignment to categories. Thus it is possible to formulate a transformation that can insert all or part of the Auxiliary Verb to the left of a Noun Phrase that precedes it, independently of what the length or internal complexity of the strings belonging to these categories may be. It is impossible, however, to formulate as a transformation such a simple operation as reflection of an arbitrary string (that is, replacement of any string $a_1 \cdots a_n$, where each a_i is a single symbol, by $a_n \cdots a_1$), or interchange of the $(2n - 1)^{\text{th}}$ word with the $2n^{\text{th}}$ word throughout a string of

arbitrary length, or insertion of a symbol in the middle of a string of even length. Similarly, if the structural analyses that define transformations are restricted to Boolean conditions on *Analyzability*, as suggested later, it will be impossible to formulate many "structure-dependent" operations as transformations — for example, an operation that will iterate a symbol that is the leftmost member of a category (impossible, short of listing all categories of the grammar in the structural analysis), or an operation that will iterate a symbol that belongs to as many rightmost as leftmost categories). Hence, one who proposes this theory would have to predict that although a language might form interrogatives, for example, by interchanging the order of certain categories (as in English), it could not form interrogatives by reflection, or interchange of odd and even words, or insertion of a marker in the middle of the sentence. Many other such predictions, none of them at all obvious in any a priori sense, can be deduced from any sufficiently explicit theory of linguistic universals that is attributed to a language-acquisition device as an intrinsic property. For some initial approaches to the very difficult but tantalizing problem of investigating questions of this sort, see Miller and Stein (1963), Miller and Norman (1964).

Notice that when we maintain that a system is not learnable by a language-acquisition device that mirrors human capacities, we do not imply that this system cannot be mastered by a human in some other way, if treated as a puzzle or intellectual exercise of some sort. The language-acquisition device is only one component of the total system of intellectual structures that can be applied to problem solving and concept formation; in other words, the *faculté de langage* is only one of the faculties of the mind. What one would expect, however, is that there should be a qualitative difference in the way in which an organism with a functional language-acquisition system⁸² will approach and deal with systems that are languagelike and others that are not.

The problem of mapping the intrinsic cognitive capacities of

an organism and identifying the systems of belief and the organization of behavior that it can readily attain should be central to experimental psychology. However, the field has not developed in this way. Learning theory has, for the most part, concentrated on what seems a much more marginal topic, namely the question of species-independent regularities in acquisition of items of a "behavioral repertoire" under experimentally manipulable conditions. Consequently, it has necessarily directed its attention to tasks that are extrinsic to an organism's cognitive capacities — tasks that must be approached in a devious, indirect, and piecemeal fashion. In the course of this work, some incidental information has been obtained about the effect of intrinsic cognitive structure and intrinsic organization of behavior on what is learned, but this has rarely been the focus of serious attention (outside of ethology). The sporadic exceptions to this observation (see, for example, the discussion of "instinctual drift" in Breland and Breland, 1961) are quite suggestive, as are many ethological studies of lower organisms. The general question and its many ramifications, however, remain in a primitive state.

In brief, it seems clear that the present situation with regard to the study of language learning is essentially as follows. We have a certain amount of evidence about the character of the generative grammars that must be the "output" of an acquisition model for language. This evidence shows clearly that taxonomic views of linguistic structure are inadequate and that knowledge of grammatical structure cannot arise by application of step-by-step inductive operations (segmentation, classification, substitution procedures, filling of slots in frames, association, etc.) of any sort that have yet been developed within linguistics, psychology, or philosophy. Further empiricist speculations contribute nothing that even faintly suggests a way of overcoming the intrinsic limitations of the methods that have so far been proposed and elaborated. In particular, such speculations have not provided any way to account for or even to express the fundamental fact about the normal use of language, namely the speaker's ability to produce and understand instantly new

sentences that are not similar to those previously heard in any physically defined sense or in terms of any notion of frames or classes of elements, nor associated with those previously heard by conditioning, nor obtainable from them by any sort of "generalization" known to psychology or philosophy. It seems plain that language acquisition is based on the child's discovery of what from a formal point of view is a deep and abstract theory—a generative grammar of his language—many of the concepts and principles of which are only remotely related to experience by long and intricate chains of unconscious quasi-inferential steps. A consideration of the character of the grammar that is acquired, the degenerate quality and narrowly limited extent of the available data, the striking uniformity of the resulting grammars, and their independence of intelligence, motivation, and emotional state, over wide ranges of variation, leave little hope that much of the structure of the language can be learned by an organism initially uninformed as to its general character.

It is, for the present, impossible to formulate an assumption about initial, innate structure rich enough to account for the fact that grammatical knowledge is attained on the basis of the evidence available to the learner. Consequently, the empiricist effort to show how the assumptions about a language-acquisition device can be *reduced to a conceptual minimum*⁸³ is quite misplaced. The real problem is that of developing a hypothesis about initial structure that is sufficiently rich to account for acquisition of language, yet not so rich as to be inconsistent with the known diversity of language. It is a matter of no concern and of only historical interest that such a hypothesis will evidently not satisfy the preconceptions about learning that derive from centuries of empiricist doctrine. These preconceptions are not only quite implausible, to begin with, but are without factual support and are hardly consistent with what little is known about how animals or humans construct a "theory of the external world."

It is clear why the view that all knowledge derives solely from the senses by elementary operations of association and "gen-

eralization" should have had much appeal in the context of eighteenth-century struggles for scientific naturalism. However, there is surely no reason today for taking seriously a position that attributes a complex human achievement entirely to months (or at most years) of experience, rather than to millions of years of evolution or to principles of neural organization that may be even more deeply grounded in physical law—a position that would, furthermore, yield the conclusion that man is, apparently, unique among animals in the way in which he acquires knowledge. Such a position is particularly implausible with regard to language, an aspect of the child's world that is a human creation and would naturally be expected to reflect intrinsic human capacity in its internal organization.

In short, the structure of particular languages may very well be largely determined by factors over which the individual has no conscious control and concerning which society may have little choice or freedom. On the basis of the best information now available, it seems reasonable to suppose that a child cannot help constructing a particular sort of transformational grammar to account for the data presented to him, any more than he can control his perception of solid objects or his attention to line and angle. Thus it may well be that the general features of language structure reflect, not so much the course of one's experience, but rather the general character of one's capacity to acquire knowledge—in the traditional sense, one's innate ideas and innate principles. It seems to me that the problem of clarifying this issue and sharpening our understanding of its many facets provides the most interesting and important reason for the study of descriptively adequate grammars and, beyond this, the formulation and justification of a general linguistic theory that meets the condition of explanatory adequacy. By pursuing this investigation, one may hope to give some real substance to the traditional belief that "the principles of grammar form an important, and very curious, part of the philosophy of the human mind" (Beattie, 1788).

§ 9. GENERATIVE CAPACITY AND ITS LINGUISTIC RELEVANCE

It may be useful to make one additional methodological observation in connection with the topics discussed in the last few sections. Given a descriptive theory of language structure,³⁴ we can distinguish its *weak generative capacity* from its *strong generative capacity* in the following way. Let us say that a grammar *weakly generates* a set of sentences and that it *strongly generates* a set of structural descriptions (recall that each structural description uniquely specifies a sentence, but not necessarily conversely), where both weak and strong generation are determined by the procedure *f* of (12iv) = (13iv) = (14iv). Suppose that the linguistic theory *T* provides the class of grammars G_1, G_2, \dots , where G_i weakly generates the language L_i and strongly generates the system of structural descriptions Σ_i . Then the class $\{L_1, L_2, \dots\}$ constitutes the *weak generative capacity* of *T* and the class $\{\Sigma_1, \Sigma_2, \dots\}$ constitutes the *strong generative capacity* of *T*.³⁵

The study of strong generative capacity is related to the study of descriptive adequacy, in the sense defined. A grammar is descriptively adequate if it strongly generates the correct set of structural descriptions. A theory is descriptively adequate if its strong generative capacity includes the system of structural descriptions for each natural language; otherwise, it is descriptively inadequate. Thus inadequacy of strong generative capacity, on empirical grounds, shows that a theory of language is seriously defective. As we have observed, however, a theory of language that appears to be empirically adequate in terms of strong generative capacity is not necessarily of any particular theoretical interest, since the crucial question of explanatory adequacy goes beyond any consideration of strong generative capacity.

The study of weak generative capacity is of rather marginal linguistic interest. It is important only in those cases where some proposed theory fails even in weak generative capacity — that is, where there is some natural language even the *sentences* of which cannot be enumerated by any grammar permitted by this theory. In fact, it has been shown that certain fairly elementary

theories (in particular, the theory of context-free phrase-structure grammar and the even weaker theory of finite-state grammar) do not have the weak generative capacity required for the description of natural language, and thus fail empirical tests of adequacy in a particularly surprising way.³⁶ From this observation we must conclude that as linguistic theory progresses to a more adequate conception of grammatical structure, it will have to permit devices with a weak generative capacity that differs, in certain respects, from that of these severely defective systems.

It is important to note, however, that the fundamental defect of these systems is not their limitation in weak generative capacity but rather their many inadequacies in strong generative capacity. Postal's demonstration that the theory of context-free grammar (simple phrase-structure grammar) fails in weak generative capacity was preceded by over a half-dozen years of discussion of the strong generative capacity of this theory, which showed conclusively that it cannot achieve descriptive adequacy. Furthermore, these limitations in strong generative capacity carry over to the theory of context-sensitive phrase-structure grammar, which probably does not fail in weak generative capacity. Presumably, discussion of weak generative capacity marks only a very early and primitive stage of the study of generative grammar. Questions of real linguistic interest arise only when strong generative capacity (descriptive adequacy) and, more important, explanatory adequacy become the focus of discussion.

As observed earlier, the critical factor in the development of a fully adequate theory is the limitation of the class of possible grammars. Clearly, this limitation must be such as to meet empirical conditions on strong (and, a fortiori, weak) generative capacity, and, furthermore, such as to permit the condition of explanatory adequacy to be met when an appropriate evaluation measure is developed. But beyond this, the problem is to impose sufficient structure on the schema that defines "generative grammar" so that relatively few hypotheses will have to be tested by the evaluation measure, given primary linguistic data. We want the hypotheses compatible with fixed data to be "scattered" in value, so that choice among them can be made relatively easily.

This requirement of "feasibility" is the major empirical constraint on a theory, once the conditions of descriptive and explanatory adequacy are met. It is important to keep the requirements of explanatory adequacy and feasibility in mind when weak and strong generative capacities of theories are studied as mathematical questions. Thus one can construct hierarchies of grammatical theories in terms of weak and strong generative capacity, but it is important to bear in mind that these hierarchies do *not* necessarily correspond to what is probably the empirically most significant dimension of increasing power of linguistic theory. This dimension is presumably to be defined in terms of the scattering in value of grammars compatible with fixed data. Along this empirically significant dimension, we should like to accept the least "powerful" theory that is empirically adequate. It might conceivably turn out that this theory is extremely powerful (perhaps even universal, that is, equivalent in generative capacity to the theory of Turing machines)³⁷ along the dimension of weak generative capacity, and even along the dimension of strong generative capacity. It will not necessarily follow that it is very powerful (and hence to be discounted) in the dimension which is ultimately of real empirical significance.

In brief, mathematical study of formal properties of grammars is, very likely, an area of linguistics of great potential. It has already provided some insight into questions of empirical interest and will perhaps some day provide much deeper insights. But it is important to realize that the questions presently being studied are primarily determined by feasibility of mathematical study, and it is important not to confuse this with the question of empirical significance.

2

Categories and Relations in Syntactic Theory

§ 1. THE SCOPE OF THE BASE

We now return to the problem of refining and elaborating the sketch (in Chapter 1, § 3) of how a generative grammar is organized. Putting off to the next chapter any question as to the adequacy of earlier accounts of grammatical transformations, we shall consider here only the formal properties of the base of the syntactic component. We are therefore concerned primarily with extremely simple sentences.

The investigation of generative grammar can profitably begin with a careful analysis of the kind of information presented in traditional grammars. Adopting this as a heuristic procedure, let us consider what a traditional grammar has to say about a simple English sentence such as the following:

(1) sincerity may frighten the boy

Concerning this sentence, a traditional grammar might provide information of the following sort:

(2) (i) the string (1) is a Sentence (S); *frighten the boy* is a Verb Phrase (VP) consisting of the Verb (V) *frighten* and the Noun Phrase (NP) *the boy*; *sincerity* is also an NP; the NP *the boy* consists of the Determiner (Det) *the*, followed by a Noun (N); the NP *sincerity* consists of just an N; *the* is, furthermore, an Article (Art); *may* is a Verbal Auxiliary (Aux) and, furthermore, a Modal (M).