

**Knowledge of  
Language:  
Its Nature, Origin, and Use**

**Noam Chomsky**

**CONVERGENCE**

**A Series Founded, Planned, and Edited  
by Ruth Nanda Anshen**

**NOTICE**

This material may be  
protected by copyright  
law (Title 17 U.S. Code.)

---

**PRAEGER**

New York  
Westport, Connecticut  
London

# CONVERGENCE

A Series Founded, Planned, and Edited  
by Ruth Nanda Anshen

## Library of Congress Cataloging in Publication Data

Chomsky, Noam.

Knowledge of language.

(Convergence)

Bibliography: p.

Includes index.

1. Language and languages. I. Title.

II. Series: Convergence (Praeger Publishers)

P106.C518 1985 410 85-12234

ISBN 0-275-90025-8

ISBN 0-275-91761-4 (pbk.)

Copyright © 1986 by Noam Chomsky

All rights reserved. No portion of this book may be reproduced, by any process or technique, without the express written consent of the publisher.

Library of Congress Catalog Card Number: 85-12234

ISBN: 0-275-91761-4

First published in 1986

Praeger Publishers, 521 Fifth Avenue, New York, NY 10175

A division of Greenwood Press, Inc.

Printed in the United States of America



The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48-1984).

10 9 8 7 6 5 4 3

## Board of Editors

Sir Fred Hoyle

Sir Bernard Lovell

Adolph Lowe

Joseph Needham

I. I. Rabi

William Richardson

Jonas Salk

Lewis Thomas

C. N. Yang

# 1 Knowledge of Language as a Focus of Inquiry

The study of language has a long and rich history, extending over thousands of years. This study has frequently been understood as an inquiry into the nature of mind and thought on the assumption that “languages are the best mirror of the human mind” (Leibniz). A common conception was that “with respect to its *substance* grammar is one and the same in all languages, though it does vary *accidentally*” (Roger Bacon). The invariant “substance” was often taken to be the mind and its acts; particular languages use various mechanisms—some rooted in human reason, others arbitrary and adventitious—for the expression of thought, which is a constant across languages. One leading eighteenth century rational grammarian defined “general grammar” as a deductive science concerned with “the immutable and general principles of spoken or written language” and their consequences; it is “prior to all languages,” because its principles “are the same as those that direct human reason in its intellectual operations” (Beauzée). Thus, “the science of language does not differ at all from the science of thought.” “Particular grammar” is not a true “science” in the sense of this rationalist tradition because it is not based solely on universal necessary laws; it is an “art” or technique that shows how given languages realize the general principles of human reason. As John Stuart Mill later expressed the same leading idea, “The principles and rules of grammar are the means by which the forms of language are made to correspond with the universal forms of thought. . . . The structure of every sentence is a lesson in logic.” Others, particularly during the

Romantic period, argued that the nature and content of thought are determined in part by the devices made available for its expression in particular languages. These devices may include contributions of individual genius that affect the "character" of a language, enriching its means of expression and the thoughts expressed without affecting its "form," its sound system and rules of word and sentence formation (Humboldt).

With regard to the acquisition of knowledge, it was widely held that the mind is not "so much to be filled therewith from without, like a vessel, as to be kindled and awaked" (Ralph Cudworth); "The growth of knowledge... [rather resembles]... the growth of Fruit; however external causes may in some degree cooperate, it is the internal vigour, and virtue of the tree, that must ripen the juices to their just maturity" (James Harris).<sup>1</sup> Applied to language, this essentially Platonistic conception would suggest that knowledge of a particular language grows and matures along a course that is in part intrinsically determined, with modifications reflecting observed usage, rather in the manner of the visual system or other bodily "organs" that develop along a course determined by genetic instructions under the triggering and shaping effects of environmental factors.

With the exception of the relativism of the Romantics, such ideas were generally regarded with much disapproval in the mainstream of linguistic research by the late nineteenth century and on through the 1950s. In part, this attitude developed under the impact of a rather narrowly construed empiricism and later behaviorist and operationalist doctrine. In part, it resulted from the quite real and impressive successes of historical and descriptive studies conducted within a narrower compass, specifically, the discovery of "sound laws" that provided much understanding of the history of languages and their relationships. In part, it was a natural consequence of the investigation of a much richer variety of languages than were known to earlier scholars, languages that appeared to violate many of the allegedly *a priori* conceptions of the earlier rationalist tradition.<sup>2</sup> After a century of general neglect or obloquy, ideas resembling those of the earlier tradition re-emerged (initially, with virtually no awareness of historical antecedents) in the mid-1950s, with the development of what came to be called "generative grammar"—again, reviving a long-lapsed and largely forgotten tradition.<sup>3</sup>

The generative grammar of a particular language (where "generative" means nothing more than "explicit") is a theory that is concerned with the form and meaning of expressions of this language. One can imagine many different kinds of approach to such questions, many points of view that might be adopted in dealing with them. Generative grammar limits itself to certain elements of this larger picture. Its standpoint is that of individual psychology. It is concerned with those aspects of form and meaning that are determined by the "language faculty," which is understood to be a particular component of the human mind. The nature of this faculty is the subject matter of a general theory of linguistic structure that aims to discover the framework of principles and elements common to attainable human languages; this theory is now often called "universal grammar" (UG), adapting a traditional term to a new context of inquiry. UG may be regarded as a characterization of the genetically determined language faculty. One may think of this faculty as a "language acquisition device," an innate component of the human mind that yields a particular language through interaction with presented experience, a device that converts experience into a system of knowledge attained: knowledge of one or another language.

The study of generative grammar represented a significant shift of focus in the approach to problems of language. Put in the simplest terms, to be elaborated below, the shift of focus was from behavior or the products of behavior to states of the mind/brain that enter into behavior. If one chooses to focus attention on this latter topic, the central concern becomes knowledge of language: its nature, origins, and use.

The three basic questions that arise, then, are these:

- (i) What constitutes knowledge of language? (1)
- (ii) How is knowledge of language acquired?
- (iii) How is knowledge of language put to use?

The answer to the first question is given by a particular generative grammar, a theory concerned with the state of the mind/brain of the person who knows a particular language. The answer to the second is given by a specification of UG along with an account of the ways in which its principles interact with experience to yield a particular language; UG is a theory of the "initial state" of the language faculty, prior to any

linguistic experience. The answer to the third question would be a theory of how the knowledge of language attained enters into the expression of thought and the understanding of presented specimens of language, and derivatively, into communication and other special uses of language.

So far, this is nothing more than the outline of a research program that takes up classical questions that had been put aside for many years. As just described, it should not be particularly controversial, since it merely expresses an interest in certain problems and offers a preliminary analysis of how they might be confronted, although as is often the case, the initial formulation of a problem may prove to be far-reaching in its implications, and ultimately controversial as it is developed.

Some elements of this picture may appear to be more controversial than they really are. Consider, for example, the idea that there is a language faculty, a component of the mind/brain that yields knowledge of language given presented experience. It is not at issue that humans attain knowledge of English, Japanese, and so forth, while rocks, birds, or apes do not under the same (or indeed any) conditions. There is, then, some property of the mind/brain that differentiates humans from rocks, birds, or apes. Is this a distinct "language faculty" with specific structure and properties, or, as some believe, is it the case that humans acquire language merely by applying generalized learning mechanisms of some sort, perhaps with greater efficiency or scope than other organisms? These are not topics for speculation or *a priori* reasoning but for empirical inquiry, and it is clear enough how to proceed: namely, by facing the questions of (1). We try to determine what is the system of knowledge that has been attained and what properties must be attributed to the initial state of the mind/brain to account for its attainment. Insofar as these properties are language-specific, either individually or in the way they are organized and composed, there is a distinct language faculty.

Generative grammar is sometimes referred to as a theory, advocated by this or that person. In fact, it is not a theory any more than chemistry is a theory. Generative grammar is a topic, which one may or may not choose to study. Of course, one can adopt a point of view from which chemistry disappears as a discipline (perhaps it is all done by angels with mirrors). In this sense, a decision to study chemistry does stake out a position on

matters of fact. Similarly, one may argue that the topic of generative grammar does not exist, although it is hard to see how to make this position minimally plausible. Within the study of generative grammar there have been many changes and differences of opinion, often reversion to ideas that had been abandoned and were later reconstructed in a different light. Evidently, this is a healthy phenomenon indicating that the discipline is alive, although it is sometimes, oddly, regarded as a serious deficiency, a sign that something is wrong with the basic approach. I will review some of these changes as we proceed.

In the mid-1950s, certain proposals were advanced as to the form that answers to the questions of (1) might take, and a research program was inaugurated to investigate the adequacy of these proposals and to sharpen and apply them. This program was one of the strands that led to the development of the cognitive sciences in the contemporary sense, sharing with other approaches the belief that certain aspects of the mind/brain can be usefully construed on the model of computational systems of rules that form and modify representations, and that are put to use in interpretation and action. From its origins (or with a longer perspective, one might say "its reincarnation") about 30 years ago, the study of generative grammar was undertaken with an eye to gaining some insight into the nature and origins of systems of knowledge, belief, and understanding more broadly, in the hope that these general questions could be illuminated by a detailed investigation of the special case of human language.

This research program has since been running its course, along a number of different paths. I will be concerned here with only one of these, with the problems it faced and the steps that were taken in an effort to deal with them. During the past 5–6 years, these efforts have converged in a somewhat unexpected way, yielding a rather different conception of the nature of language and its mental representation, one that offers interesting answers to a range of empirical questions and opens a variety of new ones to inquiry while suggesting a rethinking of the character of others. This is what accounts for an unmistakable sense of energy and anticipation—and also uncertainty—which is reminiscent of the period when the study of generative grammar in the modern sense was initiated about 30 years ago.

Some of the work now being done is quite different in character from what had previously been possible as well as considerably broader in empirical scope, and it may be that results of a rather new kind are within reach, or at least within sight. I would like to try to explain why this may be so, beginning with some remarks about goals, achievements, and failures of the past years.

To avoid misunderstanding, I am not speaking here about all of the study of language but rather of generative grammar, and even here I will not attempt anything like a real history of the course of research but rather will give a somewhat idealized picture that is in part clearer in retrospect than it was at the time. Furthermore, what I am describing has represented a minority position throughout, and probably still does, although in my view it is the correct one. A number of different current approaches share properties of the sort discussed here and may be intertranslatable to a considerable extent. I will not consider this important topic here and will also make no effort to survey the range of ideas, often conflicting, that fall within the particular tendency that I will discuss—what is now sometimes called “government-binding (GB) theory.”

I want to consider, then, two major conceptual shifts, one that inaugurated the contemporary study of generative grammar, and a second, more theory-internal, that is now in process and that offers some new perspectives on traditional problems.<sup>4</sup>

Traditional and structuralist grammar did not deal with the questions of (1), the former because of its implicit reliance on the unanalyzed intelligence of the reader, the latter because of its narrowness of scope. The concerns of traditional and generative grammar are, in a certain sense, complementary: a good traditional or pedagogical grammar provides a full list of exceptions (irregular verbs, etc.), paradigms and examples of regular constructions, and observations at various levels of detail and generality about the form and meaning of expressions. But it does not examine the question of how the reader of the grammar uses such information to attain the knowledge that is used to form and interpret new expressions, or the question of the nature and elements of this knowledge: essentially the questions of (1), above. Without too much exaggeration, one could describe such a grammar as a structured and organized version of the data presented to a child learning a language,

with some general commentary and often insightful observations. Generative grammar, in contrast, is concerned primarily with the intelligence of the reader, the principles and procedures brought to bear to attain full knowledge of a language. Structuralist theories, both in the European and American traditions, did concern themselves with analytic procedures for deriving aspects of grammar from data, as in the procedural theories of Nikolay Trubetzkoy, Zellig Harris, Bernard Bloch, and others, but primarily in the areas of phonology and morphology. The procedures suggested were seriously inadequate and in any event could not possibly be understood (and were not intended) to provide an answer to question (1ii), even in the narrower domains where most work was concentrated. Nor was there an effort to determine what was involved in offering a comprehensive account of the knowledge of the speaker/hearer.

As soon as these questions were squarely faced, a wide range of new phenomena were discovered, including quite simple ones that had passed unnoticed, and severe problems arose that had previously been ignored or seriously misunderstood. A standard belief 30 years ago was that language acquisition is a case of “overlearning.” Language was regarded as a habit system, one that was assumed to be much overdetermined by available evidence. Production and interpretation of new forms was taken to be a straightforward matter of analogy, posing no problems of principle.<sup>5</sup> Attention to the questions of (1) quickly reveals that exactly the opposite is the case: language poses in a sharp and clear form what has sometimes been called “Plato’s problem,” the problem of “poverty of stimulus,” of accounting for the richness, complexity, and specificity of shared knowledge, given the limitations of the data available. This difference of perception concerning where the problem lies—overlearning or poverty of evidence—reflects very clearly the effect of the shift of focus that inaugurated the study of generative grammar.

A great many examples have been given over the years to illustrate what clearly is the fundamental problem: the problem of poverty of evidence. A familiar example is the structure-dependence of rules, the fact that without instruction or direct evidence, children unerringly use computationally complex structure-dependent rules rather than computationally simple rules that involve only the predicate “leftmost” in a linear

sequence of words.<sup>6</sup> To take some other examples, to which we will return, consider sentences (2)–(7):

- |                                             |     |
|---------------------------------------------|-----|
| I wonder who [the men expected to see them] | (2) |
| [the men expected to see them]              | (3) |
| John ate an apple                           | (4) |
| John ate                                    | (5) |
| John is too stubborn to talk to Bill        | (6) |
| John is too stubborn to talk to             | (7) |

Both (2) and (3) include the clause bounded by brackets, but only in (2) may the pronoun *them* be referentially dependent on the antecedent *the men*; in (3) the pronoun is understood as referring in some manner indicated in the situational or discourse context, but not to the men. Numerous facts of this sort, falling under what is now generally called “binding theory,” are known without relevant experience to differentiate the cases. Such facts pose a serious problem that was not recognized in earlier work: How does every child know, unerringly, to interpret the clause differently in the two cases? And why does no pedagogic grammar have to draw the learner’s attention to such facts (which were, in fact, noticed only quite recently, in the course of the study of explicit rule systems in generative grammar)?

Turning to examples (4)–(7), sentence (5) means that John ate something or other, a fact that one might explain on the basis of a simple inductive procedure: *ate* takes an object, as in (4), and if the object is missing, it is understood as arbitrary. Applying the same inductive procedure to (6) and (7), it should be that (7) means that John is so stubborn that he (John) will not talk to some arbitrary person, on the analogy of (6). But the meaning is, in fact, quite different: namely, that John is so stubborn that some arbitrary person won’t talk to him (John). Again, this is known without training or relevant evidence.<sup>7</sup>

The situation is, in fact, more complex. Although plausible, the inductive procedure suggested for the relatively straightforward examples (4)–(5) does not seem correct. As noted by Howard Lasnik, the word *eat* has a somewhat different meaning in its intransitive usage, something like *dine*. One can say “John ate his shoe,” but “John ate” cannot be understood to include this case. The observation is general for such cases. The

intransitive forms differ from normal intransitives in other respects; for example, we can form “the dancing bear” (corresponding to “the bear that dances”), but not “the eating man” (corresponding to “the man who eats”).<sup>8</sup> Such facts pose further problems of poverty of stimulus.

Children do not make errors about the interpretation of such sentences as (6)–(7) past a certain stage of development, and if they did, the errors would largely be uncorrectable. It is doubtful that even the most compendious traditional or teaching grammar notes such simple facts as those illustrated in (2)–(7), and such observations lie far beyond the domain of structural grammars. A wide variety of examples of this sort immediately come to attention when one faces the questions formulated in (1).

Knowledge of language is often characterized as a practical ability to speak and understand, so that questions (Ii) and (Iiii) are closely related, perhaps identified. Ordinary usage makes a much sharper distinction between the two questions, and is right to do so. Two people may share exactly the same knowledge of language but differ markedly in their ability to put this knowledge to use. Ability to use language may improve or decline without any change in knowledge. This ability may also be impaired, selectively or in general, with no loss of knowledge, a fact that would become clear if injury leading to impairment recedes and lost ability is recovered. Many such considerations support the commonsense assumption that knowledge cannot be properly described as a practical ability. Furthermore, even if this view could somehow be maintained, it would leave open all of the serious questions. Thus, what is the nature of the “practical ability” manifested in our interpretation of the sentences (2)–(7), how is it properly described, and how is it acquired?

Often it is not immediately obvious what our knowledge of language entails in particular cases, a fact illustrated even with short and simple sentences such as (8)–(10):

- |                                               |      |
|-----------------------------------------------|------|
| his wife loves her husband                    | (8)  |
| John is too clever to expect us to catch Bill | (9)  |
| John is too clever to expect us to catch      | (10) |

In the case of (8), it takes some thought to determine whether *his* can be referentially dependent on *her husband* if *her* is

dependent on *his wife*—that is, if the reference of either *he* or *she* is not somehow contextually indicated.<sup>9</sup> Examples (9) and (10) are, in fact, analogous to (6) and (7), respectively, but again, it takes some thought to discover that (10) means that John is so clever that an arbitrary person cannot expect us to catch him (John), although it is clear at once that it does not mean that John is so clever that he (John) cannot catch some arbitrary person, on the analogy of (9) (and (4), (5)). Our abilities seem limited somehow in such cases (and there are far more complex ones), but it would make little sense to speak of our knowledge of language as “limited” in any comparable way.

Suppose we insist on speaking of knowledge of language as a practical ability to speak and understand. Then normal usage must be revised in numerous cases such as those just discussed. Suppose that Jones takes a public speaking course and improves his ability to speak and understand without any change in his knowledge of English, as we would describe the situation in normal usage. We must now revise this common-sense usage and say, rather, that Jones has improved his ability<sub>1</sub> to use his ability<sub>2</sub> to speak and understand; similar translations are required in the other cases. But the two occurrences of “ability” in this description are hardly more than homonyms. Ability<sub>1</sub> is ability in the normal sense of the word: it can improve or decline, can be inadequate to determine consequences of knowledge, and so on. Ability<sub>2</sub>, however, remains stable while our ability to use it changes, and we have this kind of “ability” even when we are unable to detect what it entails in concrete cases. In short, the neologism “ability<sub>2</sub>” is invested with all the properties of knowledge. Note that there are cases when we do speak of abilities that we cannot put to use: for example, the case of swimmers who cannot swim because their hands are tied, although they retain the ability to swim. The cases in question are not of this sort, however.

The purpose of the attempt to reduce knowledge to ability is, presumably, to avoid problematic features that seem to inhere in the concept of knowledge, to show that these can be explained in dispositional or other terms more closely related to actual behavior (whether this is possible even in the case of ability<sub>1</sub>, the normal sense, is another question). But nothing of the sort is achieved by this departure from ordinary usage; the

problems remain, exactly as before, now embedded in terminological confusion. The task of determining the nature of our knowledge (= ability<sub>2</sub>), and accounting for its origins and use, remains exactly as challenging as before, despite the terminological innovations.

Other examples similar to (8)–(10) raise further questions. Consider the following sentences:

John is too stubborn to expect anyone to talk to Bill (11)

John is too stubborn to visit anyone who talked to Bill (12)

Suppose we delete *Bill* from (11) and (12), yielding (13) and (14), respectively:

John is too stubborn to expect anyone to talk to (13)

John is too stubborn to visit anyone who talked to (14)

Sentence (13) is structurally analogous to (10), and is understood in the same manner: it means that John is so stubborn that an arbitrary person would not expect anyone to talk to him (John). “By analogy,” then, we would expect sentence (14) to mean that John is so stubborn that an arbitrary person would not visit anyone who talked to him (John). But it does not have that meaning; in fact, it is gibberish. Here we have a double failure of analogy. Sentence (14) is not understood “on the analogy” of (4), (5), (6), (9), and (12) (hence meaning that John is so stubborn that he (John) would not visit anyone who talked to some arbitrary person), nor is it understood “on the analogy” of (7), (10), and (13); rather, it has no interpretation at all. And while the status of (11), (12), and (14) is immediately obvious, it takes some thought or preparation to see that (13) has the interpretation it does have, and thus to determine the consequences of our knowledge in this case.

Again, these are facts that we know, however difficult it may be to determine that our system of knowledge has these consequences. We know these facts without instruction or even

direct evidence, surely without correction of error by the speech community. It would be absurd to try to teach such facts as these to people learning English as a second language, just as no one taught them to us or even presented us with evidence that could yield this knowledge by any generally reliable procedure. This is knowledge without grounds, without good reasons or support by reliable procedures in any general or otherwise useful sense of these notions. Were we to insist that knowledge is a kind of ability, we would have to claim that we lack the ability to understand "John is too stubborn to talk to" as meaning "John is too stubborn to talk to someone or other" (on the analogy of "John ate an apple" — "John ate"), and that we lack the ability to understand (14) on the analogy of "John ate an apple" — "John ate" (so that it means that John is too stubborn to visit anyone who talked to someone or other) or on the analogy of "John is too stubborn to talk to," with the "inversion strategy" that we somehow use in this case (so that (14) means that John is too stubborn for someone or other to visit anyone who talked to him, John). But these would be odd claims, to say the least. These are not failures of ability. It is not that we are too weak, or lack some special skill that could be acquired. We are perfectly capable of associating the sentence (14), for example, with either of the two meanings that would be provided "by analogy" (or others), but we know that these are not the associations that our knowledge of the language provides; ability is one thing, knowledge something quite different. The system of knowledge that has somehow developed in our minds has certain consequences, not others; it relates sound and meaning and assigns structural properties to physical events in certain ways, not others.

It seems that there is little hope in accounting for our knowledge in terms of such ideas as analogy, induction, association, reliable procedures, good reasons, and justification in any generally useful sense, or in terms of "generalized learning mechanisms" (if such exist). And it seems that we should follow normal usage in distinguishing clearly between knowledge and ability to use that knowledge. We should, so it appears, think of knowledge of language as a certain state of the mind/brain, a relatively stable element in transitory mental states once it is attained; furthermore, as a state of some distinguishable

faculty of the mind—the language faculty—with its specific properties, structure, and organization, one "module" of the mind.<sup>10</sup>

## NOTES

1. On these and many other discussions, primarily in the seventeenth–nineteenth centuries, see Chomsky (1966). For discussion of some misinterpretation of this work, see Bracken (1984).

2. The alleged *a priorism* of work in this tradition has often been exaggerated. See Chomsky (1966) and more recent work for discussion of this point.

3. The tradition, in this case, is a different one, represented in its most advanced form in the early work of the Indian grammarians 2,500 years ago. See Kiparsky (1982). A modern counterpart is Bloomfield (1939), which was radically different in character from the work of the period and inconsistent with his own theories of language, and remained virtually without influence or even awareness despite Bloomfield's great prestige.

4. See Newmeyer (1980) for one view of the history of this period prior to the second major conceptual shift; and for some more personal comments, the introduction to Chomsky (1975a), a somewhat abbreviated version of a 1956 revision of a 1955 manuscript, both unpublished. See Lightfoot (1982) and Hornstein and Lightfoot (1981) for discussion of the general backgrounds for much current work, and Radford (1981) for an introduction to the work that led to the second conceptual shift. See Chomsky (1981) for a more technical presentation of some of the ideas that entered into this conceptual shift and van Riemsdijk and Williams (1985) for an introductory study of this current work.

5. Although basically adopting this point of view, W.V. Quine, however, argued that there is a very severe, in fact, insuperable problem of underdetermination affecting all aspects of language and grammar, and much of psychology more generally (Quine, 1960, 1972). I do not think that he succeeded in showing that some novel form of indeterminacy affects the study of language beyond the normal underdetermination of theory by evidence; his own formulations of the thesis furthermore involve internal inconsistency (see Chomsky, 1975b, 1980b). There seems no reason on these grounds, then, to distinguish linguistics or psychology in principle from the natural sciences in accordance with what Hockney (1975) calls Quine's "bifurcation

thesis." A similar conclusion is reached by Putnam (1981) in his abandonment of metaphysical realism on Quinean grounds. His step also abandons the bifurcation thesis, although in the opposite direction.

6. See Chomsky (1975a). See Crain and Nakayama (1984) for empirical study of this question with 3–5-year-old children.

7. The reaction to such phenomena, also unnoticed until recently, again illustrates the difference of outlook of structuralist-descriptive and generative grammar. For some practitioners of the former, the statement of the facts, which is straightforward enough once they are observed, is the answer—nothing else is necessary; for the latter, the statement of the facts poses the problem to be solved. Cf. Ney (1983), particularly, his puzzlement about the "peculiar view of grammar [that] unnecessarily complicates the whole matter" by seeking an explanation for the facts. Note that there is no question of right or wrong here, but rather of topic of inquiry.

8. In early work, such facts were used to motivate an analysis of intransitives such as *eat* as derived from corresponding transitives by a system of ordered rules that excluded the unwanted cases; see Chomsky (1962).

9. On structures of this type, and problems of binding theory, more generally, see Higginbotham (1988a), among much other work.

10. See Fodor (1983). But it is too narrow to regard the "language module" as an input system in Fodor's sense, if only because it is used in speaking and thought. We might consider supplementing this picture by adding an "output system," but plainly this must be linked to the input system; we do not expect a person to speak only English and understand only Japanese. That is, the input and output systems must each access a fixed system of knowledge. The latter, however, is a central system which has essential problems of modularity, a fact that brings the entire picture into question. Furthermore, even regarded as an input system, the language module does not appear to have the property of rapidity of access that Fodor discusses, as indicated by (8)–(14). Note also that even if Fodor is right in believing that there is a sharp distinction between modules in his sense and "the rest," which is holistic in several respects, it does not follow that the residue is unstructured. In fact, this seems highly unlikely, if only because of the "epistemic boundedness" that he notes. Many other questions arise concerning Fodor's very intriguing discussion of these issues, which I will not pursue here.

## 2

## Concepts of Language

### 2.1 THE COMMONSENSE CONCEPT AND DEPARTURES FROM IT

Let us turn now to the questions of (1) of Chapter 1. To begin with, let us distinguish the intuitive, pretheoretic commonsense notion of language from various technical concepts that have been proposed with the intent of developing an eventual science of language. Let us call the latter "scientific approaches" to language, with an eye directed more toward a possible future than a present reality, some might argue. The scientific approaches, I believe without exception, depart from the commonsense notion in several ways; these departures also affect the concepts of knowledge or understanding of language, use of language, rule of language, rule-guided linguistic behavior, and others.

In the first place, the commonsense notion of language has a crucial sociopolitical dimension. We speak of Chinese as "a language," although the various "Chinese dialects" are as diverse as the several Romance languages. We speak of Dutch and German as two separate languages, although some dialects of German are very close to dialects that we call "Dutch" and are not mutually intelligible with others that we call "German." A standard remark in introductory linguistics courses is that a language is a dialect with an army and a navy (attributed to Max Weinreich). That any coherent account can be given of "language" in this sense is doubtful; surely, none has been offered or even seriously attempted. Rather, all scientific approaches have simply abandoned these elements of what is called "language" in common usage.<sup>1</sup>

The commonsense notion also has a normative-teleological element that is eliminated from scientific approaches. I do not refer here to prescriptive grammar but to something else. Consider the way we describe a child or a foreigner learning English. We have no way of referring directly to what that person knows: It is not English, nor is it some other language that resembles English. We do not, for example, say that the person has a perfect knowledge of some language L, similar to English but still different from it. What we say is that the child or foreigner has a "partial knowledge of English," or is "on his or her way" toward acquiring knowledge of English, and if they reach the goal, they will then know English. Whether or not a coherent account can be given of this aspect of the commonsense terminology, it does not seem to be one that has any role in an eventual science of language.

I will follow standard practice in disregarding these aspects of the commonsense notions of language and the associated notions of rule-following and so forth, although the departure should be noted, and one may ask whether it is entirely innocent.

Modern linguistics commonly avoided these questions by considering an idealized "speech community" that is internally consistent in its linguistic practice.<sup>2</sup> For Leonard Bloomfield, for example, a language is "the totality of utterances that can be made in a speech community," regarded as homogeneous (Bloomfield, 1928/1957). In other scientific approaches, the same assumption enters in one or another form, explicitly or tacitly, in identification of the object of inquiry. No attempt is made to capture or formulate any concept with the sociopolitical or normative-teleological aspects of informal usage of the term "language." The same is true of approaches that understand language to be a social product in accordance with the Saussurean concept of "langue."

Of course, it is understood that speech communities in the Bloomfieldian sense—that is, collections of individuals with the same speech behavior<sup>3</sup>—do not exist in the real world. Each individual has acquired a language in the course of complex social interactions with people who vary in the ways in which they speak and interpret what they hear and in the internal representations that underlie their use of language. Structural linguistics abstracted from these facts in its attempts at theory construction; we also abstract from these facts in posing ques-

tions (1) of Chapter 1, considering only the case of a person presented with uniform experience in an ideal Bloomfieldian speech community with no dialect diversity and no variation among speakers.

We should also make note of a more subtle theory-internal assumption: The language of the hypothesized speech community, apart from being uniform, is taken to be a "pure" instance of UG in a sense that must be made precise, and to which we will return. We exclude, for example, a speech community of uniform speakers, each of whom speaks a mixture of Russian and French (say, an idealized version of the nineteenth-century Russian aristocracy). The language of such a speech community would not be "pure" in the relevant sense, because it would not represent a single set of choices among the options permitted by UG but rather would include "contradictory" choices for certain of these options.

Questions (1) of Chapter 1, then, arise initially under these idealizations, and the same is true, in effect, of other approaches to language, although the fact is often not explicitly recognized and may even sometimes be denied.

The legitimacy of these idealizations has sometimes been questioned, but on dubious grounds.<sup>4</sup> Indeed, they seem indispensable. Surely there is some property of mind P that would enable a person to acquire a language under conditions of pure and uniform experience, and surely P (characterized by UG) is put to use under the real conditions of language acquisition. To deny these assumptions would be bizarre indeed: It would be to claim either that language can be learned only under conditions of diversity and conflicting evidence, which is absurd, or that the property P exists—there exists a capacity to learn language in the pure and uniform case—but the actual learning of language does not involve this capacity. In the latter case, we would ask why P exists; is it a "vestigial organ" of some sort? The natural approach, and one that I think is tacitly adopted even by those who deny the fact, is to attempt to determine the real property of mind P, and then ask how P functions under the more complex conditions of actual linguistic diversity. It seems clear that any reasonable study of the nature, acquisition, and use of language in real life circumstances must accept these assumptions and then proceed on the basis of some tentative characterization of the property of mind P. In short, the ideali-

zations made explicit in more careful work are hardly controversial; they isolate for examination a property of the language faculty the existence of which is hardly in doubt, and which is surely a crucial element in actual language acquisition.

By making these idealizations explicit and pursuing our inquiry in accordance with them, we do not in any way prejudice the study of language as a social product. On the contrary, it is difficult to imagine how such studies might fruitfully progress without taking into account the real properties of mind that enter into the acquisition of language, specifically, the properties of the initial state of the language faculty characterized by UG.

Note also that the study of language and UG, conducted within the framework of individual psychology, allows for the possibility that the state of knowledge attained may itself include some kind of reference to the social nature of language. Consider, for example, what Putnam (1975) has called "the division of linguistic labor." In the language of a given individual, many words are semantically indeterminate in a special sense: The person will defer to "experts" to sharpen or fix their reference. Suppose, for example, that someone knows that yawls and ketches are sailing vessels but is unsure of the exact reference of the words "yawl" and "ketch," leaving it to specialists to fix this reference. In the lexicon of this person's language, the entries for "yawl" and "ketch" will be specified to the extent of his or her knowledge, with an indication that details are to be filled in by others, an idea that can be made precise in various ways but without going beyond the study of the system of knowledge of language of a particular individual. Other social aspects of language can be regarded in a like manner—although this is not to deny the possibility or value of other kinds of study of language that incorporate social structure and interaction. Contrary to what is sometimes thought, no conflicts of principle or practice arise in this connection.

We are also assuming another idealization: That the property of mind described by UG is a species characteristic, common to all humans. We thus abstract from possible variation among humans in the language faculty. It is plausible to suppose that apart from pathology (potentially an important area of inquiry), such variation as there may be is marginal and can be safely ignored across a broad range of linguistic investigation. Again,

in scientific approaches. Weaker assumptions than strict identity would suffice for the discussion below, but this stronger assumption seems a reasonable one, to a very good approximation, and I will keep to it here.

## 2.2 EXTERNALIZED LANGUAGE

Scientific approaches to language, in the sense of the term used earlier, have developed various technical notions of language to replace the commonsense notion. The term "grammar" has also been used in a variety of ways. In conventional usage, a grammar is a description or theory of a language, an object constructed by a linguist. Let us keep to this usage. Then associated with the various technical notions of language there are corresponding notions of grammar and of universal grammar (UG).

Structural and descriptive linguistics, behavioral psychology, and other contemporary approaches tended to view a language as a collection of actions, or utterances, or linguistic forms (words, sentences) paired with meanings, or as a system of linguistic forms or events. In Saussurean structuralism, a language (*langue*) was taken to be a system of sounds and an associated system of concepts; the notion of sentence was left in a kind of limbo, perhaps to be accommodated within the study of language use. For Bloomfield, as noted earlier, a language is "the totality of utterances that can be made in a speech community." The American variety of structural-descriptive linguistics that was heavily influenced by Bloomfield's ideas furthermore concentrated primarily on sound and word structure, apart from various proposals, notably those of Zellig Harris, as to how larger units (phrases) could be constructed by analytic principles modelled on those introduced for phonology and morphology.<sup>5</sup> Many researchers today adopt a position of the sort lucidly developed by David Lewis, who defines a language as a pairing of sentences and meanings (the latter taken to be set-theoretic constructions in terms of possible worlds) over an infinite range, where the language is "used by a population" when certain regularities "in action or belief" hold among the population with reference to the language, sustained by an interest in communication.<sup>6</sup>

Let us refer to such technical concepts as instances of "externalized language" (E-language), in the sense that the construct is understood independently of the properties of the mind/brain. Under the same rubric we may include the notion of language as a collection (or system) of actions or behaviors of some sort. From a point of view such as this, a grammar is a collection of descriptive statements concerning the E-language, the actual or potential speech events (perhaps along with some account of their context of use or semantic content). In technical terms, the grammar may be regarded as a function that enumerates the elements of the E-language. Sometimes, grammar has been regarded as a property of E-language, as in Bloomfield's remark that a grammar is "the meaningful arrangement of forms in a language" (Bloomfield, 1933). Despite appearances, the problem of accounting for the unbounded character of the E-language and the person's knowledge of language including this fundamental property is not squarely addressed in such approaches, a matter to which we will return.

The E-language is now understood to be the real object of study. Grammar is a derivative notion; the linguist is free to select the grammar one way or another as long as it correctly identifies the E-language. Apart from this consideration, questions of truth and falsity do not arise. Quine, for example, has argued that it is senseless to take one grammar rather than another to be "correct" if they are extensionally equivalent, characterizing the same E-language, for him a set of expressions (Quine, 1972). And Lewis doubts that there is any way "to make objective sense of the assertion that a grammar *G* is used by a population *P* whereas another grammar *G'*, which generates the same language as *G*, is not."

The notion of E-language is familiar from the study of formal systems, as is the conclusion just cited: In the case of the "language of arithmetic," for example, there is no objective sense to the idea that one set of rules that generates the well-formed formulas is correct and another wrong.

As for UG, to the extent that such a study was recognized as legitimate, this theory would consist of statements that are true of many or all human languages, perhaps a set of conditions satisfied by the E-languages that count as human languages. Some appeared to deny the possibility of the enterprise, for example, Martin Joos, who put forth what he called the "Boas-

ian" view that "languages could differ from each other without limit and in unpredictable ways," echoing William Dwight Whitney's reference to "the infinite diversity of human speech" and Edward Sapir's notion that "language is a human activity that varies without assignable limit."<sup>7</sup> Such statements reflect a fairly broad consensus of the time. Although they could hardly have been intended literally, they did express a relativistic impulse that denigrated the study of UG. More precisely, it cannot be that human language varies without assignable limit, although it might be true that it is "infinitely diverse"; it is an empirical question of some interest whether UG permits an infinite variety of possible languages (or a variety that is infinite in more than structurally trivial respects, say, with no bound on vocabulary), or only a finite diversity.<sup>8</sup>

Nevertheless, significant contributions were made to UG in our sense within these traditions. For example, the theory of distinctive features in phonology, which greatly influenced structuralist studies in other fields, postulated a fixed inventory of "atomic elements" from which phonological systems could be drawn, with certain general laws and implicational relations governing the choice. And it was generally assumed that such notions as topic and comment, or subject and predicate, were universal features of language, reflecting the fact that a declarative sentence is about something and says something about it. Later, important work on linguistic universals was conducted by Joseph Greenberg and others, yielding many generalizations that require explanation, for example, the fact that if a language has subject-object-verb order, it will tend to have postpositions rather than prepositions, and so on.

Along these lines, then, we may develop a certain technical concept of language (E-language), and an associated concept of grammar and UG, as a basis for a scientific study of language. Many different specific ideas fall roughly within this general framework.

### 2.3 INTERNALIZED LANGUAGE

A rather different approach was taken, for example, by Otto Jespersen, who held that there is some "notion of structure" in the mind of the speaker "which is definite enough to guide

him in framing sentences of his own," in particular, "free expressions" that may be new to the speaker and to others.<sup>9</sup> Let us refer to this "notion of structure" as an "internalized language" (I-language). The I-language, then, is some element of the mind of the person who knows the language, acquired by the learner, and used by the speaker-hearer.

Taking language to be I-language, the grammar would then be a theory of the I-language, which is the object under investigation. And if, indeed, such a "notion of structure" exists, as Jespersen held, then questions of truth and falsity arise for grammar as they do for any scientific theory. This way of approaching the questions of language is radically different from the one sketched above and leads to a very different conception of the nature of the inquiry.

Let us return now to the point of view outlined in Chapter 1. Knowing the language *L* is a property of a person *H*; one task of the brain sciences is to determine what it is about *H*'s brain by virtue of which this property holds. We suggested that for *H* to know the language *L* is for *H*'s mind/brain to be in a certain state; more narrowly, for the language faculty, one module of this system, to be in a certain state  $S_L$ .<sup>10</sup> One task of the brain sciences, then, is to discover the mechanisms that are the physical realization of the state  $S_L$ .

Suppose we analyze the notion "H knows language *L*" in relational terms, that is, as involving a relation *R* (knowing, having, or whatever) holding between *H* and an abstract entity *L*. One might question this move; we speak of a person as knowing U.S. history without assuming that there is an entity, U.S. history, that the person knows, or knows in part. Let us, however, assume the move to be legitimate in this case. The assumption will be justified to the extent that this move contributes to providing insight into the questions that primarily concern us, those of (1) of Chapter 1; this would be the case, for example, if there are significant principles governing the set of postulated entities *L*. Suppose that we proceed further to regard talk of mind as talk about the brain undertaken at a certain level of abstraction at which we believe, rightly or wrongly, that significant properties and explanatory principles can be discovered. Then statements about *R* and *L* belong to the theory of mind, and one task of the brain sciences will be to explain what

it is about *H*'s brain (in particular, its language faculty) that corresponds to *H*'s knowing *L*, that is, by virtue of which  $R(H, L)$  holds and the statement that  $R(H, L)$  is true.

It is natural to take *L* to be I-language, Jespersen's "notion of structure," regarding this as an entity abstracted from a state of the language faculty, the latter being one component of the mind. Then, for *H* to know *L* is for *H* to have a certain I-language. The statements of a grammar are statements of the theory of mind about the I-language, hence statements about structures of the brain formulated at a certain level of abstraction from mechanisms. These structures are specific things in the world, with their specific properties. The statements of a grammar or the statement that  $R(H, L)$  are similar to statements of a physical theory that characterizes certain entities and their properties in abstraction from whatever may turn out to be the mechanisms that account for these properties: say, a nineteenth-century theory about valence or properties expressed in the periodic table. Statements about I-language or the statement that  $R(H, L)$  (for various choices of *H* and *L*) are true or false, much in the way that statements about the chemical structure of benzene, or about the valence of oxygen, or about chlorine and fluorine being in the same column of the periodic table are true or false. The I-language *L* may be the one used by a speaker but not the I-language *L'*, even if the two generate the same class of expressions (or other formal objects) in whatever precise sense we give to this derivative notion; *L'* may not even be a possible human I-language, one attainable by the language faculty.

UG now is construed as the theory of human I-languages, a system of conditions deriving from the human biological endowment that identifies the I-languages that are humanly accessible under normal conditions. These are the I-languages *L* such that  $R(H, L)$  may be true (for normal *H*, under normal conditions).<sup>11</sup>

Of course, there is no guarantee that this way of approaching the problems of (1) in Chapter 1 is the correct one. This approach may turn out to be thoroughly misguided, even if it achieves substantial success—just as a theory of valence, etc. might have turned out to be completely off the track, despite its substantial success in nineteenth-century chemistry. It is always reasonable to consider alternative approaches, if they can be

devised, and this will remain true no matter what successes are achieved. The situation does not seem different in principle from what we find in other areas of empirical inquiry. I will suggest directly that in certain fundamental respects early ideas about I-language were misguided and should be replaced by a rather different conception, although one formulated in the same general framework. The reasons, however, do not derive from any incoherence or flaw in the general approach but rather from empirical considerations of description and explanation.

## 2.4 THE SHIFT OF FOCUS FROM E-LANGUAGE TO I-LANGUAGE

### 2.4.1 On the Reasons for the Shift of Focus

In Chapter 1, we saw that the study of generative grammar shifted the focus of attention from actual or potential behavior and the products of behavior to the system of knowledge that underlies the use and understanding of language, and more deeply, to the innate endowment that makes it possible for humans to attain such knowledge. The shift in focus was from the study of E-language to the study of I-language, from the study of language regarded as an externalized object to the study of the system of knowledge of language attained and internally represented in the mind/brain. A generative grammar is not a set of statements about externalized objects constructed in some manner. Rather, it purports to depict exactly what one knows when one knows a language: that is, what has been learned, as supplemented by innate principles. UG is a characterization of these innate, biologically determined principles, which constitute one component of the human mind—the language faculty.

With this shift of focus, we at once face the questions (1) of Chapter 1. In the earliest work, the answer to (1i) was taken to be that knowledge of language is knowledge of a certain rule system; the answer to (1ii), that this knowledge arises from an initial state  $S_0$  that converts experience to a “steady state”  $S_s$ , which incorporates an I-language. Acquisition of language is,

then, a matter of adding to one's store of rules, or modifying this system, as new data are processed. Question (1iii) breaks down into two parts: a “perception problem” and a “production problem.” The perception problem would be dealt with by construction of a parser that incorporates the rules of the I-language along with other elements: a certain organization of memory and access (perhaps a deterministic pushdown structure with buffer of a certain size; see Marcus, 1980), certain heuristics, and so forth. A parser should not map expressions into their structures in the way that these are associated by the I-language. For example, a parser should fail to do so in the case of so-called “garden-path sentences”<sup>12</sup> or sentences that overload memory for left-to-right pass, it should mirror the difficulties experienced with sentences such as (8)–(14) of Chapter 1 and so forth. The production problem is considerably more obscure; we will return to that.

The E-language that was the object of study in most of traditional or structuralist grammar or behavioral psychology is now regarded as an epiphenomenon at best. Its status is similar to that of other derivative objects, say, the set of rhyming pairs, which is also determined by the I-language that constitutes the system of knowledge attained. One might argue that the status of the E-language is considerably more obscure than that of the set of rhyming pairs, since the latter is determined in a fairly definite way by the I-language whereas the bounds of E-language can be set one way or another, depending on some rather arbitrary decisions as to what it should include.

Summarizing, then, we have the following general picture. The language faculty is a distinct system of the mind/brain, with an initial state  $S_0$  common to the species (to a very close first approximation, apart from pathology, etc.) and apparently unique to it in essential respects.<sup>13</sup> Given appropriate experience, this faculty passes from the state  $S_0$  to some relatively stable steady state  $S_s$ , which then undergoes only peripheral modification (say, acquiring new vocabulary items). The attained state incorporates an I-language (it is the state of having or knowing a particular I-language). UG is the theory of  $S_s$ ; particular grammars are theories of various I-languages. The I-languages that can be attained with  $S_0$  fixed and experience varying are the attainable human languages, where by “language” we now

In this respect, the study of language as understood in the discussion above is like chemistry, biology, solar physics, or the theory of human vision. Whether the burden of proof faced by advocates of A-linguistics can be borne, I will not speculate, except to observe that even if it can, the fact would have no consequences with regard to the legitimacy or character of the enterprise we are discussing, as Soames makes clear.

Note that the issue is not the legitimacy of abstraction. It is perfectly proper to develop the subject of rational mechanics, a branch of mathematics abstracted from physics that treats planets as mass points obeying certain laws, or to develop theories that consider aspects of I-language in abstraction from their physical realization or other properties; indeed, that is the standard practice, as outlined earlier. But one is not misled thereby into believing that the subject matter of rational mechanics is an entity in a Platonic heaven, and there is no more reason to suppose that that is true in the study of language.<sup>23</sup>

#### 2.4.2 The Empirical Basis for the Study of I-language

In actual practice, linguistics as a discipline is characterized by attention to certain kinds of evidence that are, for the moment, readily accessible and informative: largely, the judgments of native speakers. Each such judgment is, in fact, the result of an experiment, one that is poorly designed but rich in the evidence it provides. In practice, we tend to operate on the assumption, or pretense, that these informant judgments give us "direct evidence" as to the structure of the I-language, but, of course, this is only a tentative and inexact working hypothesis, and any skilled practitioner has at his or her disposal an armory of techniques to help compensate for the errors introduced. In general, informant judgments do not reflect the structure of the language directly; judgments of acceptability, for example, may fail to provide direct evidence as to grammatical status because of the intrusion of numerous other factors. The same is true of other judgments concerning form and meaning. These are, or should be, truisms.<sup>24</sup>

In principle, evidence concerning the character of the I-language and initial state could come from many different

sources apart from judgments concerning the form and meaning of expressions: perceptual experiments, the study of acquisition and deficit or of partially invented languages such as creoles,<sup>25</sup> or of literary usage or language change, neurology, biochemistry, and so on. It was one of the many contributions of the late Roman Jakobson to have emphasized this fact, in principle, and in his own work in practice. As in the case of any inquiry into some aspect of the physical world, there is no way of delimiting the kinds of evidence that might, in principle, prove relevant. The study of language structure as currently practiced should eventually disappear as a discipline as new types of evidence become available, remaining distinct only insofar as its concern is a particular faculty of the mind, ultimately the brain: its initial state and its various attainable mature states.

To be sure, the judgments of native speakers will always provide relevant evidence for the study of language, just as perceptual judgments will always provide relevant evidence for the study of human vision, although one would hope that such evidence will eventually lose its uniquely privileged status. If a theory of language failed to account for these judgments, it would plainly be a failure; we might, in fact, conclude that it is not a theory of language, but rather of something else. But we cannot know in advance just how informative various kinds of evidence will prove to be with regard to the language faculty and its manifestations, and we should anticipate that a broader range of evidence and deeper understanding will enable us to identify in just what respects informant judgments are useful or unreliable and why, and to compensate for the errors introduced under the tentative working assumption, which is indispensable, for today, and does provide us with rich and significant information.

It is important to bear in mind that the study of one language may provide crucial evidence concerning the structure of some other language, if we continue to accept the plausible assumption that the capacity to acquire language, the subject matter of UG, is common across the species. This conclusion is implicit in the research program outlined earlier. A study of English is a study of the realization of the initial state  $S_0$  under particular conditions. Therefore, it embodies assumptions, which should be made explicit, concerning  $S_0$ . But  $S_0$  is a

constant; therefore, Japanese must be an instantiation of the same initial state under different conditions. Investigation of Japanese might show that the assumptions concerning  $S_0$  derived from the study of English were incorrect; these assumptions might provide the wrong answers for Japanese, and after correcting them on this basis we might be led to modify the postulated grammar of English. Because evidence from Japanese can evidently bear on the correctness of a theory of  $S_0$ , it can have indirect—but very powerful—bearing on the choice of the grammar that attempts to characterize the I-language attained by a speaker of English. This is standard practice in the study of generative grammar. For this reason alone it is quite wrong to suppose that there are no grounds to choose among “extensionally equivalent grammars” for a “given language” (see pp. 20, 30–1): One of these might, for example, require a theory of  $S_0$  that is demonstrably inadequate for some other language.

On the highly relativistic assumptions of certain varieties of descriptive linguistics that held that each language must be studied in its own terms, this research program may seem to be senseless or illegitimate, although one should note that this point of view was, in part, an ideology that was not observed in practice. If we are interested in discovering the real properties of the initial state of the language faculty and of its particular realizations as potential or actual I-languages, the ideology must be abandoned, and we must regard a theory of one language as subject to change on the basis of evidence concerning other languages (mediated through a theory of UG), or evidence of other sorts.

We observed that it is a task for the brain sciences to explain the properties and principles discovered in the study of mind. More accurately, the interdependency of the brain sciences and the study of mind is reciprocal. The theory of mind aims to determine the properties of the initial state  $S_0$  and each attainable state  $S_L$  of the language faculty, and the brain sciences seek to discover the mechanisms of the brain that are the physical realizations of these states. There is a common enterprise: to discover the correct characterization of the language faculty in its initial and attained states, to discover the truth about the language faculty. This enterprise is conducted at several levels: an abstract characterization in the theory of mind, and an

inquiry into mechanisms in the brain sciences. In principle, discoveries about the brain should influence the theory of mind, and at the same time the abstract study of states of the language faculty should formulate properties to be explained by the theory of the brain and is likely to be indispensable in the search for mechanisms. To the extent that such connections can be established, the study of the mind—in particular, of I-language—will be assimilated to the mainstream of the natural sciences.

So little is now known about the relevant aspects of the brain that we can barely even speculate about what the connections might be. We can, however, imagine how they might be established in principle, however remote the goal. Suppose that the study of I-language establishes certain general principles of binding theory that explain facts of the sort discussed in Chapter 1. Then a task of the brain sciences is to determine what mechanisms are responsible for the fact that these principles hold. Suppose that we have two grammars—two theories of the state of knowledge attained by a particular person—and suppose further that these theories are “extensionally equivalent” in the sense that they determine the same E-language in whatever sense we give to this derivative notion. It could in principle turn out that one of these grammars incorporates properties and principles that are readily explained in terms of brain mechanisms whereas the other does not. Similarly, two theories of UG that are equivalent in that they specify exactly the same set of attainable I-languages might be distinguishable in terms of properties of the brain. For example, one might contain certain principles and possibilities of variation that can be readily explained in terms of brain mechanisms, and the other not.

It is easy enough to imagine cases of this sort. Suppose that theory I contains the principles  $P_1, \dots, P_n$  and theory II contains the principles  $Q_1, \dots, Q_m$ , and that the two theories are logically equivalent: The principles of each can be deduced from the principles of the other so that any description of behavior or potential behavior in terms of one of these theories can be reformulated in terms of the other. It could be that the brain sciences would show that each  $P_i$  corresponds to some determinate complex of neural mechanisms, whereas there is no such

account of the  $Q_i$ 's; some brain injury, for example, might selectively modify the  $P_i$ 's but not the  $Q_i$ 's. In such a case, facts about the brain would select among theories of the mind that might be empirically indistinguishable in other terms. Although results of this sort are remote in the current state of understanding, they are possible. The relation of brain and mind, so conceived, is a problem of the natural sciences.

### 2.4.3 Some Consequences of the Shift of Focus

To summarize, we may think of a person's knowledge of a particular language as a state of the mind, realized in some arrangement of physical mechanisms. We abstract the I-language as "what is known" by a person in this state of knowledge. This finite system, the I-language, is what the linguist's generative grammar attempts to characterize. If I say that this system has such-and-such properties, what I say is true or false. I am, in short, proposing a theoretical account of the properties of certain mechanisms, an account presented at a level of abstraction at which we believe that significant properties of these mechanisms can be expressed and principles governing these mechanisms and their functions elucidated. The study is in some ways similar to what Gunther Stent has called "cerebral hermeneutics," referring to the abstract investigation of the ways in which the visual system constructs and interprets visual experience (Stent, 1981). Similarly, UG is the study of one aspect of biological endowment, analogous to the study of the innate principles that determine that we will have a human rather than an insect visual system. The technical concept "knowledge of I-language" is a reasonably close approximation to what is informally called "knowledge of language," abstracting from several aspects of the commonsense notion as discussed earlier, although this consideration is a secondary one for reasons already mentioned.

The shift of point of view to a mentalist interpretation of the study of language was, as noted earlier, one factor in the development of the contemporary cognitive sciences, and constituted a step toward the incorporation of the study of language within the natural sciences, because it helps pave the way to an inquiry into the mechanisms with the properties exhibited in

the study of rules and representations. This shift also led at once to a recasting of many of the traditional questions of language study. Many new and challenging problems arose, while a number of familiar problems dissolved when viewed from this perspective.

Consider the study of sound structure, the primary focus of attention in structural and descriptive linguistics. Taking English as the topic of inquiry, the problem is to discover the elements into which the stream of speech is subdivided and their properties and structural arrangements: phonemes and features, regarded as segments of an acoustic wave form or of a series of articulatory motions. Much of phonological theory consisted of analytic procedures for accomplishing this task. Focusing on the I-language, however, the problem is a rather different one: to find the mental representations that underlie the production and perception of speech and the rules that relate these representations to the physical events of speech. The problem is to find the best theory to account for a wide variety of facts, and we do not expect that analytic procedures exist to accomplish this task, just as there are no such procedures in other fields.

Consider, for example, the words listed below, where column I is the conventional orthography, column II appears

I	II	III
bet	bet	bet
bent	bent	bēt
bend	bend	bend
knot	nat	nat
nod	nɒd	nAd
write	rayt	rayt
ride	rayd	rAyd
writer	rayt+r	rayDr
rider	rayd+r	rAydDr

to be the correct phonological representation, and column III, the approximate phonetic representations in one dialect of English, taking [a] to be a short vowel and [A] a corresponding long vowel (their exact phonetic character is irrelevant here),

To proceed, we are now concerned with I-language and the initial state of the language faculty, with the linguist's grammars and UG. As a tentative empirical hypothesis, we might take the I-language to be a rule system of some sort, a specific realization of the options permitted by UG, fixed by presented experience. The rule system assigns to each expression a structure, which we may take to be a set of representations, one on each linguistic level, where a linguistic level is a particular system of mental representation. This structure must provide whatever information about an expression is available to the person who knows the language, insofar as this information derives from the language faculty; its representations must specify just what the language faculty contributes to determining how the expression is produced, used, and understood.

A linguistic level is a system consisting of a set of minimal elements (primes), an operation of concatenation that forms strings of primes, as much mathematical apparatus as is necessary to construct appropriate formal objects from these elements, the relevant relations that hold of these elements, and a class of designated formal objects (markers) that are assigned to expressions as their representations on this level. The rule system expresses the relations among the various levels in the language in question and determines the elements and properties of each level. At the level of phrase structure, for example, the primes are the minimal elements that enter into syntactic description (*John*, *run*, *past-tense*, N, V, S, etc.), the basic relation is *is-a* (*John* is an N, *John ran* is an S, etc.), and the phrase-markers will be certain formal objects constructed out of primes that express completely the relation *is-a*. The phrase-marker for the string *John ran* will indicate that the full string is an S (sentence), that *John* is an N (noun) and an NP (noun phrase), and that *ran* is a V (verb) and a VP (verb phrase); examples appear below.

The theory of linguistic structure (UG) will have the task of specifying these concepts precisely.<sup>28</sup> The theory must provide grammars for the I-languages that can, in principle, be attained by a human mind/brain, given appropriate experience,<sup>29</sup> and it must furthermore be so constrained that just the right I-language is determined, given the kind of evidence that suffices for language acquisition. We turn next to these questions.

## NOTES

1. These observations, generally considered truisms, are rejected by Katz (1981, pp. 79–80) on the grounds that to recognize the fact that the concepts language and dialect of colloquial usage involve a sociopolitical dimension would be "like claiming that the concept of number is not a concept of mathematics but a sociopolitical one." There is no reason to accept this curious conclusion.
2. However, there were exceptions, for example, the theory of "overall patterns," of which each English dialect was held to be a subsystem. See Trager and Smith (1951). Note that the question of "variable rules," as discussed by some sociolinguists, is not relevant here.
3. We put aside here just what this term would mean in Bloomfieldian or any other variety of "behaviorist" linguistics. Pursuing such an approach, one would have to explain just what it means to say that people speak the very same language although they do not tend to say the same things in given circumstances. The same question arises if language is defined as a "complex of present dispositions to verbal behavior" (Quine, 1960), as do other problems that seem insoluble if the technical constructed concept "language" is to be a useful term for the investigation of language, or to have any relation to what we call "language." On this matter, see Chomsky (1975b, pp. 192–195).
4. One might also note some unintentionally comical objections, such as the charge by Oxford professor of linguistics Roy Harris (1983) that the standard idealization (which he ascribes to Saussure-Bloomfield-Chomsky) reflects "a fascist concept of language if ever there was one," because it takes the "ideal" speech community to be "totally homogeneous."
5. For some discussion, see Chomsky (1964) and Postal (1964). For comparison of transformational generative grammar with Harris's early theory of transformations, regarded as an analytic procedure applying beyond the sentence level of "structural grammar," see the introduction to Chomsky (1975a).
6. Lewis (1975). Lewis provides one of the clearest presentations of an "extensional" approach to language and also a critique of studies of "internalized language" in the sense described below. For critical discussion, see Chomsky (1980b).
7. Editorial comments in Joos (1957); Whinney (1872); Sapir (1921). Whinney, who exerted a major influence on Saussure and American linguistics, was criticizing Steinthal's Humboldtian approach, which I believe, falls naturally into the earlier tradition referred to

above. Humboldt, who is widely regarded (e.g., by Bloomfield) as an extreme relativist, in fact held that "all languages with regard to their grammar are very similar, if they are investigated not superficially, but deeply in their inner nature." See Chomsky (1966), p. 90, and references cited, for further discussion.

8. This question, however, was surely not what Whitney had in mind.

9. Jespersen (1924). On Jespersen's notions as compared to those of contemporary generative grammar, see Reynolds (1971); Chomsky (1977), Chapter 1.

10. One might argue that the systems we are considering constitute only one element of the faculty of language, understood more broadly to encompass other capacities involved in the use and understanding of language, for example, what is sometimes called "communicative competence," or parts of the human conceptual system that are specifically related to language. See Chomsky (1980b). I will put such questions aside here, continuing to use the term "language faculty" in the narrower sense of the previous discussion.

11. For a related but somewhat different way of viewing these questions, see Higginbotham (1983b).

12. Those that tend to yield a false parse, such as Thomas Bever's example "the horse raced past the barn fell," where the first six words are generally taken to constitute a full clause, leaving no interpretation for the final word, although on reflection it is clear that the expression is a well-formed sentence stating that a certain horse fell, namely, the one that was raced past the barn.

13. Obviously, the questions of innateness and species-specificity are distinct. It has been alleged that I and others have taken "innate" and "species-specific" to be "synonyms" (Cartmill, 1984). I am unaware of any examples of such confusion, although there are a number of articles relaying it.

14. See Chomsky (1980b), pp. 134-139.

15. On this matter, see Marr (1982). Note that the question of the legitimacy or sense of a realist interpretation of science in general is not at issue here; rather, nothing new in principle seems to arise in the case of the study of I-language and its origins. If one wants to consider the question of realism, psychology and linguistics seem poor choices; the question should be raised with regard to the more advanced sciences, where there is much better hope of gaining insight into the matter. See Chomsky (1980b) for further discussion.

16. For some commentary on the general issue, see Enc (1983).

17. On some misunderstandings, which are repeated in subsequent work that I will not discuss here, see Chomsky (1980b), pp.

123-128. As for the publishing history, the earliest publications on generative grammar were presented in a framework suggested by certain topics in automata theory (e.g., my *Syntactic Structures*, 1957—actually course notes for an undergraduate course at MIT and hence presented from a point of view related to interests of these students). Specifically linguistic work, such as Chomsky (1975a), was not publishable at the time. In the latter, considerations of weak generative capacity (i.e., characterizability of E-languages), finite automata and the like were completely absent, and emphasis was on I-language, although the term was not used.

18. For further discussion of this matter, see Chomsky (1980b).

19. See Katz (1981) and Bever (1983).

20. This would follow if the evidence stipulated to be relevant to identifying a certain Platonic language as P-English is distinct from the evidence that bears on the theory of the I-language actually represented in the mind/brains of speakers of English, or if some novel canons are adopted for interpreting evidence. By a similar procedure, we could establish "Platonic biology," concerned, for example, with what Katz calls the "essential property" of a heart (that it is a pump) and thus abstracting from the physical laws that make it beat (a nonessential property). We might then find that the best biological theory is distinct from the best theory of Platonic biology just as the best (ultimately, biological) theory of I-language might be distinct from the best theory of Platonic language (however it is specified; for Katz, by analysis of "our concept of the abstract object natural language").

21. For some recent discussion of the matter in connection with historical linguistics, see Lightfoot (1979).

22. Katz insists that disciplines such as chemistry, biology, and so forth have inherent, conceptually determined boundaries. Indeed, he regards the claim as uncontroversial, the alternative being a form of "nihilism" that "would turn the spectrum of well-focused academic disciplines into chaos" (*op. cit.*).

23. Arguments that have been offered to the contrary seem to me question-begging or otherwise flawed. Thus, Katz argues against Hilary Putnam that if what we call "cats" were discovered to be robots controlled from outer space, then they would not be cats, because the meaning of "cat" in the Platonic entity P-English is "feline animal"; this would remain true even if it were determined that in the I-language of each speaker of English, "cat" is understood in accordance with Putnam's analysis, which takes cats to be of the same natural kind (a concept of science) as particular exemplars. The argument goes through, trivially, with regard to P-English as Katz stipulates its

properties. But Putnam was proposing a theory concerning human languages and conceptual systems, concerning English, not P-English as Katz defines it, and Katz offers no reason to believe that his Platonic object merits the name "English" any more than an equally legitimate abstract object that would incorporate Putnam's assumptions. Throughout, the arguments are of this sort. Katz also presents an account of the history of generative grammar and of documents he cites that is seriously inaccurate, as is often evident even on internal grounds. See also Chomsky (1981), pp. 314-315.

24. For discussion of some common misunderstandings about these and related matters, see Newmeyer (1983).

25. On the relevance of this material, see Bickerton (1984) and references cited, and discussion in the same issue of the journal.

26. For some discussion, see Chomsky (1980b, 1981); and Chomsky, Huybregts, and van Riemsdijk (1982).

27. The matter is more complex. See Evans (1980) and Higgins-botham (1983a). But we can put aside the required sharpening of these notions here.

28. For an early effort, see Chomsky (1975a), dating from 1955-56.

29. A stronger requirement would be that UG specify *exactly* the I-languages attainable under normal conditions. It is not obvious, however, that UG meets this conditions. The attainable languages are those that fall in the intersection of those determined by UG and the humanly learnable systems, and conditions on learnability might exclude certain grammars permitted by UG. Similar remarks hold with regard to parsing. For background on these matters, see Wexler and Culicover (1980) and Berwick and Weinberg (1984).

### 3 Facing Plato's Problem

#### 3.1. A MODEL OF EXPLANATION

With the shift of focus from E-language to I-language, from the study of behavior and its products to the study of systems of mental representation and computation, a number of questions arise. Some relate to the legitimacy or the proper bounds of this move; these I will put aside until Chapter 4. Others arise internally to the research program that develops naturally from this shift of focus. These are substantive questions as to how the general ideas should be fleshed out. Let us now turn to these questions.

The central task is to find the basic elements of I-language—henceforth, language. One must, in the first place, show that the devices provided by the theory of UG are adequate to the descriptive task at hand—i.e., that they are rich enough to account for the attested variety of languages and, indeed, for their possible variety. A second task is to show that these devices are meager enough so that very few languages are made available to the language-learner, given data that, in fact, suffice for language acquisition. Unless this condition is satisfied by the theory of UG, it will be impossible to account for the fact that languages are learned. The transition from the initial state to the steady state takes place in a determinate fashion, with no conscious attention or choice. The transition is essentially uniform for individuals in a given speech community despite diverse experience. The state attained is highly articulated and very rich, providing a specific interpretation for a vast array of sentences lacking close models in our experience. These conditions of our variant of Plato's problem set the second task that the research program must address.