1 Linguistic contributions to the study of mind: past

In these lectures, I would like to focus attention on the question, What contribution can the study of language make to our understanding of human nature? In one or another manifestation, this question threads its way through modern Western thought. In an age that was less self-conscious and less compartmentalized than ours, the nature of language, the respects in which language mirrors human mental processes or shapes the flow and character of thought – these were topics for study and speculation by scholars and gifted amateurs with a wide variety of interests, points of view, and intellectual backgrounds. And in the nineteenth and twentieth centuries, as linguistics, philosophy, and psychology have uneasily tried to go their separate ways, the classical problems of language and mind have inevitably reappeared and have served to link these diverging fields and to give direction and significance to their efforts. There have been signs in the past decade that the rather artificial separation of disciplines may be coming to an end. It is no longer a point of honor for each to demonstrate its absolute independence of the others, and new interests have emerged that permit the classical problems to be formulated in novel and occasionally suggestive ways – for example, in terms of the new perspectives provided by cybernetics and the communication sciences, and against the background of developments in comparative and physiological psychology that challenge long-standing convictions and free the scientific imagination from certain shackles that had become so familiar a part of our intellectual environment as to be almost beyond awareness. All of this is highly encouraging. I think there is more of a healthy ferment in cognitive psychology – and in the particular branch of cognitive psychology known as linguistics – than there has been for many years. And one of the most encouraging signs is that skepticism with regard to the orthodoxies of the recent past is coupled with an awareness of the temptations and the dangers of premature orthodoxy, an awareness that, if it can persist, may prevent the rise of new and stultifying dogma.

It is easy to be misled in an assessment of the current scene; nevertheless, it seems to me that the decline of dogmatism and the accompanying search for new approaches to old and often still intractable problems are quite unmistakable, not only in linguistics but in all of the disciplines concerned with the

study of mind. I remember quite clearly my own feeling of uneasiness as a student at the fact that, so it seemed, the basic problems of the field were solved, and that what remained was to sharpen and improve techniques of linguistic analysis that were reasonably well understood and to apply them to a wider range of linguistic materials. In the postwar years, this was a dominant attitude in most active centers of research. I recall being told by a distinguished anthropological linguist, in 1953, that he had no intention of working through a vast collection of materials that he had assembled because within a few years it would surely be possible to program a computer to construct a grammar from a large corpus of data by the use of techniques that were already fairly well formalized. At the time, this did not seem an unreasonable attitude, though the prospect was saddening for anyone who felt, or at least hoped, that the resources of human intelligence were somewhat deeper than these procedures and techniques might reveal. Correspondingly, there was a striking decline in studies of linguistic method in the early 1950s as the most active theoretical minds turned to the problem of how an essentially closed body of technique could be applied to some new domain – say, to analysis of connected discourse, or to other cultural phenomena beyond language. I arrived at Harvard as a graduate student shortly after B. F. Skinner had delivered his William James Lectures, later to be published in his book *Verbal Behavior.* Among those active in research in the philosophy or psychology of language, there was then little doubt that although details were missing, and although matters could not really be quite that simple, nevertheless a behavioristic framework of the sort Skinner had outlined would prove quite adequate to accommodate the full range of language use. There was now little reason to question the conviction of Leonard Bloomfield, Bertrand Russell, and positivistic linguists, psychologists, and philosophers in general that the framework of stimulus-response psychology would soon be extended to the point where it would provide a satisfying explanation for the most mysterious of human abilities. The most radical souls felt that perhaps, in order to do full justice to these abilities, one must postulate little *s*'s and *r*'s inside the brain alongside the capital *S*'s and *R*'s that were open to immediate inspection, but this extension was not inconsistent with the general picture.

Critical voices, even those that commanded considerable prestige, were simply unheard. For example, Karl Lashley gave a brilliant critique of the prevailing framework of ideas in 1948, arguing that underlying language use – and all organized behavior – there must be abstract mechanisms of some sort that are not analyzable in terms of association and that could not have been developed by any such simple means. But his arguments and proposals, though sound and perceptive, had absolutely no effect on the development of the field and went by unnoticed even at his own university (Harvard), then the leading center of psycholinguistic research. Ten years later Lashley's contribution began to be

appreciated, but only after his insights had been independently achieved in another context.

The technological advances of the 1940s simply reinforced the general euphoria. Computers were on the horizon, and their imminent availability reinforced the belief that it would suffice to gain a theoretical understanding of only the simplest and most superficially obvious of phenomena – everything else would merely prove to be "more of the same," an apparent complexity that would be disentangled by the electronic marvels. The sound spectrograph, developed during the war, offered similar promise for the physical analysis of speech sounds. The interdisciplinary conferences on speech analysis of the early 1950s make interesting reading today. There were few so benighted as to question the possibility, in fact the immediacy, of a final solution to the problem of converting speech into writing by available engineering technique. And just a few years later, it was jubilantly discovered that machine translation and automatic abstracting were also just around the corner. For those who sought a more mathematical formulation of the basic processes, there was the newly developed mathematical theory of communication, which, it was widely believed in the early 1950s, had provided a fundamental concept – the concept of "information" – that would unify the social and behavioral sciences and permit the development of a solid and satisfactory mathematical theory of human behavior on a probabilistic base. At about the same time, the theory of automata developed as an independent study, making use of closely related mathematical notions. And it was linked at once, and quite properly, to earlier explorations of the theory of neural nets. There were those – John von Neumann, for example – who felt that the entire development was dubious and shaky at best, and probably quite misconceived, but such qualms did not go far to dispel the feeling that mathematics, technology, and behavioristic linguistics and psychology were converging on a point of view that was very simple, very clear, and fully adequate to provide a basic understanding of what tradition had left shrouded in mystery.

In the United States at least, there is little trace today of the illusions of the early postwar years. If we consider the current status of structural linguistic methodology, stimulus-response psycholinguistics (whether or not extended to "mediation theory"), or probabilistic or automata-theoretic models for language use, we find that in each case a parallel development has taken place: a careful analysis has shown that insofar as the system of concepts and principles that was advanced can be made precise, it can be demonstrated to be inadequate in a fundamental way. The kinds of structures that are realizable in terms of these theories are simply not those that must be postulated to underlie the use of language, if empirical conditions of adequacy are to be satisfied. What is more, the character of the failure and inadequacy is such as to give little reason to believe that these approaches are on the right track. That is, in each case it

has been argued – quite persuasively, in my opinion – that the approach is not only inadequate but misguided in basic and important ways. It has, I believe, become quite clear that if we are ever to understand how language is used or acquired, then we must abstract for separate and independent study a cognitive system, a system of knowledge and belief, that develops in early childhood and that interacts with many other factors to determine the kinds of behavior that we observe; to introduce a technical term, we must isolate and study the system of *linguistic competence* that underlies behavior but that is not realized in any direct or simple way in behavior. And this system of linguistic competence is qualitatively different from anything that can be described in terms of the taxonomic methods of structural linguistics, the concepts of S-R psychology, or the notions developed within the mathematical theory of communication or the theory of simple automata. The theories and models that were developed to describe simple and immediately given phenomena cannot incorporate the real system of linguistic competence; "extrapolation" for simple descriptions cannot approach the reality of linguistic competence; mental structures are not simply "more of the same" but are qualitatively different from the complex networks and structures that can be developed by elaboration of the concepts that seemed so promising to many scientists just a few years ago. What is involved is not a matter of degree of complexity but rather a quality of complexity. Correspondingly, there is no reason to expect that the available technology can provide significant insight or understanding or useful achievements; it has noticeably failed to do so, and, in fact, an appreciable investment of time, energy, and money in the use of computers for linguistic research – appreciable by the standards of a small field like linguistics – has not provided any significant advance in our understanding of the use or nature of language. These judgments are harsh, but I think they are defensible. They are, furthermore, hardly debated by active linguistic or psycholinguistic researchers.

At the same time there have been significant advances, I believe, in our understanding of the nature of linguistic competence and some of the ways in which it is put to use, but these advances, such as they are, have proceeded from assumptions very different from those that were so enthusiastically put forth in the period I have been discussing. What is more, these advances have not narrowed the gap between what is known and what can be seen to lie beyond the scope of present understanding and technique; rather, each advance has made it clear that these intellectual horizons are far more remote than was heretofore imagined. Finally, it has become fairly clear, it seems to me, that the assumptions and approaches that appear to be productive today have a distinctly traditional flavor to them; in general, a much despised tradition has been largely revitalized in recent years and its contributions given some serious and, I believe, well-deserved attention. From the recognition of these facts flows the general and quite healthy attitude of skepticism that I spoke of earlier.

In short, it seems to me quite appropriate, at this moment in the development of linguistics and psychology in general, to turn again to classical questions and to ask what new insights have been achieved that bear on them, and how the classical issues may provide direction for contemporary research and study.

When we turn to the history of study and speculation concerning the nature of mind and, more specifically, the nature of human language, our attention quite naturally comes to focus on the seventeenth century, "the century of genius," in which the foundations of modern science were firmly established and the problems that still confound us were formulated with remarkable clarity and perspicuity. There are many far from superficial respects in which the intellectual climate of today resembles that of seventeenth-century Western Europe. One, particularly crucial in the present context, is the very great interest in the potentialities and capacities of automata, a problem that intrigued the seventeenth-century mind as fully as it does our own. I mentioned above that there is a slowly dawning realization that a significant gap – more accurately, a yawning chasm – separates the system of concepts of which we have a fairly clear grasp, on the one hand, and the nature of human intelligence, on the other. A similar realization lies at the base of Cartesian philosophy. Descartes also arrived, quite early in his investigations, at the conclusion that the study of mind faces us with a problem of quality of complexity, not merely degree of complexity. He felt that he had demonstrated that understanding and will, the two fundamental properties of the human mind, involved capacities and principles that are not realizable by even the most complex of automata.

It is particularly interesting to trace the development of this argument in the works of the minor and now quite forgotten Cartesian philosophers, like Cordemoy, who wrote a fascinating treatise extending Descartes' few remarks about language, or La Forge, who produced a long and detailed *Traité de l'esprit de l'homme* expressing, so he claimed with some reason, what Descartes would likely have said about this subject had he lived to extend his theory of man beyond physiology. One may question the details of this argument, and one can show how it was impeded and distorted by certain remnants of scholastic doctrine – the framework of substance and mode, for example. But the general structure of the argument is not unreasonable; it is, in fact, rather analogous to the argument against the framework of ideas of the early postwar years, which I mentioned at the outset of this lecture. The Cartesians tried to show that when the theory of corporeal body is sharpened and clarified and extended to its limits, it is still incapable of accounting for facts that are obvious to introspection and that are also confirmed by our observation of the actions of other humans. In particular, it cannot account for the normal use of human language, just as it cannot explain the basic properties of thought. Consequently, it becomes necessary to invoke an entirely new principle – in Cartesian terms, to postulate a second substance whose essence is thought, alongside of body,

with its essential properties of extension and motion. This new principle has a "creative aspect," which is evidenced most clearly in what we may refer to as "the creative aspect of language use," the distinctively human ability to express new thoughts and to understand entirely new expressions of thought, within the framework of an "instituted language," a language that is a cultural product subject to laws and principles partially unique to it and partially reflections of general properties of mind. These laws and principles, it is maintained, are not formulable in terms of even the most elaborate extension of the concepts proper to the analysis of behavior and interaction of physical bodies, and they are not realizable by even the most complex automaton. In fact, Descartes argued that the only sure indication that another body possesses a human mind, that it is not a mere automaton, is its ability to use language in the normal way; and he argued that this ability cannot be detected in an animal or an automaton which, in other respects, shows signs of apparent intelligence exceeding those of a human, even though such an organism or machine might be as fully endowed as a human with the physiological organs necessary to produce speech.

I will return to this argument and the ways in which it was developed. But I think it is important to stress that, with all its gaps and deficiencies, it is an argument that must be taken seriously. There is nothing at all absurd in the conclusion. It seems to me quite possible that at that particular moment in the development of Western thought there was the possibility for the birth of a science of psychology of a sort that still does not exist, a psychology that begins with the problem of characterizing various systems of human knowledge and belief, the concepts in terms of which they are organized and the principles that underlie them, and that only then turns to the study of how these systems might have developed through some combination of innate structure and organism – environment interaction. Such a psychology would contrast rather sharply with the approach to human intelligence that begins by postulating, on a priori grounds, certain specific mechanisms that, it is claimed, *must* be those underlying the acquisition of all knowledge and belief. The distinction is one to which I will return in a subsequent lecture. For the moment, I want merely to stress the reasonableness of the rejected alternative and, what is more, its consistency with the approach that proved so successful in the seventeenth-century revolution in physics.

There are methodological parallels that have perhaps been inadequately appreciated between the Cartesian postulation of a substance whose essence was thought and the post-Newtonian acceptance of a principle of attraction as an innate property of the ultimate corpuscles of matter, an active principle that governs the motions of bodies. Perhaps the most far-reaching contribution of Cartesian philosophy to modern thought was its rejection of the scholastic notion of substantial forms and real qualities, of all those "little images fluttering

through the air" to which Descartes referred with derision. With the exorcism of these occult qualities, the stage was set for the rise of a physics of matter in motion and a psychology that explored the properties of mind. But Newton argued that Descartes' mechanical physics wouldn't work – the second book of the *Principia* is largely devoted to this demonstration – and that it is necessary to postulate a new force to account for the motion of bodies. The postulate of an attractive force acting at a distance was inconsistent with the clear and distinct ideas of common sense and could not be tolerated by an orthodox Cartesian – such a force was merely another occult quality. Newton quite agreed, and he attempted repeatedly to find a mechanical explanation of the cause of gravity. He rejected the view that gravity is "essential and inherent to matter" and maintained that "to tell us that every species of things is endowed with an occult specific property (such as gravity) by which it acts and produces manifest effects, is to tell us nothing." Some historians of science have suggested that Newton hoped, like Descartes, to write a *Principles of Philosophy* but that his failure to explain the cause of gravity on mechanical grounds restricted him to a *Mathematical Principles of Natural Philosophy.* Thus, to the common sense of Newton as well as the Cartesians, physics was still not adequately grounded, because it postulated a mystical force capable of action at a distance. Similarly, Descartes' postulation of mind as an explanatory principle was unacceptable to the empiricist temper. But the astonishing success of mathematical physics carried the day against these common-sense objections, and the prestige of the new physics was so high that the speculative psychology of the Enlightenment took for granted the necessity of working within the Newtonian framework, rather than on the Newtonian analogy – a very different matter. The occult force of gravity was accepted as an obvious element of the physical world, requiring no explanation, and it became inconceivable that one might have to postulate entirely new principles of functioning and organization outside the framework of what soon became the new "common sense." Partly for this reason, the search for an analogous scientific psychology that would explore the principles of mind, whatever they might be, was not undertaken with the thoroughness that was then, as now, quite possible.

I do not want to overlook a fundamental distinction between the postulation of gravity and the postulation of a *res cogitans,* namely the enormous disparity in the power of the explanatory theories that were developed. Nevertheless, I think it is instructive to note that the reasons for the dissatisfaction of Newton, Leibnitz, and the orthodox Cartesians with the new physics are strikingly similar to the grounds on which a dualistic rationalist psychology was soon to be rejected. I think it is correct to say that the study of properties and organization of mind was prematurely abandoned, in part on quite spurious grounds, and also to point out that there is a certain irony in the common view that its abandonment was caused by the gradual spread of a more general "scientific" attitude.

I have tried to call attention to some similarities between the intellectual climate of the seventeenth century and that of today. It is illuminating, I think, to trace in somewhat greater detail the specific course of development of linguistic theory during the modern period, in the context of the study of mind and of behavior in general.[1]

A good place to begin is with the writings of the Spanish physician Juan Huarte, who in the late sixteenth century published a widely translated study on the nature of human intelligence. In the course of his investigations, Huarte came to wonder at the fact that the word for "intelligence," *ingenio*, seems to have the same Latin root as various words meaning "engender" or "generate." This, he argued, gives a clue to the nature of mind. Thus, "One may discern two generative powers in man, one common with the beasts and the plants, and the other participating of spiritual substance. Wit (Ingenio) is a generative power. The understanding is a generative faculty." Huarte's etymology is actually not very good; the insight, however, is quite substantial.

Huarte goes on to distinguish three levels of intelligence. The lowest of these is the "docile wit," which satisfies the maxim that he, along with Leibnitz and many others, wrongly attributes to Aristotle, namely that there is nothing in the mind that is not simply transmitted to it by the senses. The next higher level, normal human intelligence, goes well beyond the empiricist limitation: it is able to "engender within itself, by its own power, the principles on which knowledge rests." Normal human minds are such that "assisted by the subject alone, without the help of anybody, they will produce a thousand conceits they never heard spoke of . . . inventing and saying such things as they never heard from their masters, nor any mouth." Thus, normal human intelligence is capable of acquiring knowledge through its own internal resources, perhaps making use of the data of sense but going on to construct a cognitive system in terms of concepts and principles that are developed on independent grounds; and it is capable of generating new thoughts and of finding appropriate and novel ways of expressing them, in ways that entirely transcend any training or experience.

Huarte postulates a third kind of wit, "by means of which some, without art or study, speak such subtle and surprising things, yet true, that were never before seen, heard, or writ, no, nor ever so much as thought of." The reference here is to true creativity, an exercise of the creative imagination in ways that go beyond normal intelligence and may, he felt, involve "a mixture of madness."

Huarte maintains that the distinction between docile wit, which meets the empiricist maxim, and normal intelligence, with its full generative capacities, is the distinction between beast and man. As a physician, Huarte was much

[1] For additional details and discussion, see my *Cartesian Linguistics* (New York: Harper & Row, 1966) and the references cited there.

interested in pathology. In particular, he notes that the most severe disability of wit that can afflict a human is a restriction to the lowest of the three levels, to the docile wit that conforms to empiricist principles. This disability, says Huarte, "resembles that of Eunuchs, incapable of generation." Under these sad circumstances, in which the intelligence can only receive stimuli transmitted by sense and associate them with one another, true education is of course impossible, since the ideas and principles that permit the growth of knowledge and understanding are lacking. In this case, then, "neither the lash of the rod, nor cries, nor method, nor examples, nor time, nor experience, nor anything in nature can sufficiently excite him to bring forth anything."

Huarte's framework is useful for discussing "psychological theory" in the ensuing period. Typical of later thought is his reference to use of language as an index of human intelligence, of what distinguishes man from animals, and, specifically, his emphasis on the creative capacity of normal intelligence. These concerns dominated rationalist psychology and linguistics. With the rise of romanticism, attention shifted to the third type of wit, to true creativity, although the rationalist assumption that normal human intelligence is uniquely free and creative and beyond the bounds of mechanical explanation was not abandoned and played an important role in the psychology of romanticism, and even in its social philosophy.

As I have already mentioned, the rationalist theory of language, which was to prove extremely rich in insight and achievement, developed in part out of a concern with the problem of other minds. A fair amount of effort was devoted to a consideration of the ability of animals to follow spoken commands, to express their emotional states, to communicate with one another, and even apparently to cooperate for a common goal; all of this, it was argued, could be accounted for on "mechanical grounds," as this notion was then understood – that is, through the functioning of physiological mechanisms in terms of which one could formulate the properties of reflexes, conditioning and reinforcement, association, and so on. Animals do not lack appropriate organs of communication, nor are they simply lower along some scale of "general intelligence."

In fact, as Descartes himself quite correctly observed, language is a species-specific human possession, and even at low levels of intelligence, at pathological levels, we find a command of language that is totally unattainable by an ape that may, in other respects, surpass a human imbecile in problem-solving ability and other adaptive behavior. I will return later to the status of this observation, in the light of what is now known about animal communication. There is a basic element lacking in animals, Descartes argued, as it is lacking in even the most complex automaton that develops its "intellectual structures" completely in terms of conditioning and association – namely Huarte's second type of wit, the generative ability that is revealed in the normal human use of language as a free instrument of thought. If by experiment we convince ourselves that

another organism gives evidence of the normal, creative use of language, we must suppose that it, like us, has a mind and that what it does lies beyond the bounds of mechanical explanation, outside the framework of the stimulus-response psychology of the time, which in relevant essentials is not significantly different from that of today, though it falls short in sharpness of technique and scope and reliability of information.

It should not be thought, incidentally, that the only Cartesian arguments for the beast-machine hypothesis were those derived from the apparent inability of animals to manifest the creative aspect of language use. There were also many others – for example, the natural fear of population explosion in the domains of the spirit if every gnat had a soul. Or the argument of Cardinal Melchior de Polignac, who argued that the beast-machine hypothesis followed from the assumption of the goodness of God, since, as he pointed out, one can see "how much more humane is the doctrine that animals suffer no pain."[2] Or there is the argument of Louis Racine, son of the dramatist, who was struck by the following insight: "If beasts had souls and were capable of feelings, would they show themselves insensible to the affront and injustice done them by Descartes? Would they not rather have risen up in wrath against the leader and the sect which so degraded them?" One should add, I suppose, that Louis Racine was regarded by his contemporaries as the living proof that a brilliant father could not have a brilliant son. But the fact is that the discussion of the existence of other minds, and, in contrast, the mechanical nature of animals, continually returned to the creative aspect of language use, to the claim that – as formulated by another minor seventeenth-century figure – "if beasts reasoned, they would be capable of true speech with its infinite variety."

It is important to understand just what properties of language were most striking to Descartes and his followers. The discussion of what I have been calling "the creative aspect of language use" turns on three important observations. The first is that the normal use of language is innovative, in the sense that much of what we say in the course of normal language use is entirely new, not a repetition of anything that we have heard before and not even similar in pattern – in any useful sense of the terms "similar" and "pattern" – to sentences or discourse that we have heard in the past. This is a truism, but an important one, often overlooked and not infrequently denied in the behaviorist period of linguistics to which I referred earlier, when it was almost universally claimed that a person's knowledge of language is representable as a stored set of patterns, overlearned through constant repetition and detailed training, with innovation being at most a matter of "analogy." The fact surely is, however, that the number of sentences

[2] These examples are taken from the excellent study by Leonora Cohen Rosenfield, *From Beast-Machine to Man-Machine* (New York: Oxford University Press, 1941). The quotes are her paraphrases of the original.

in one's native language that one will immediately understand with no feeling of difficulty or strangeness is astronomical; and that the number of patterns underlying our normal use of language and corresponding to meaningful and easily comprehensible sentences in our language is orders of magnitude greater than the number of seconds in a lifetime. It is in this sense that the normal use of language is innovative.

However, in the Cartesian view even animal behavior is potentially infinite in its variety, in the special sense in which the readings of a speedometer can be said, with an obvious idealization, to be potentially infinite in variety. That is, if animal behavior is controlled by external stimuli or internal states (the latter including those established by conditioning), then as the stimuli vary over an indefinite range, so may the behavior of the animal. But the normal use of language is not only innovative and potentially infinite in scope, but also free from the control of detectable stimuli, either external or internal. It is because of this freedom from stimulus control that language can serve as an instrument of thought and self-expression, as it does not only for the exceptionally gifted and talented, but also, in fact, for every normal human.

Still, the properties of being unbounded and free from stimulus control do not, in themselves, exceed the bounds of mechanical explanation. And Cartesian discussion of the limits of mechanical explanation therefore took note of a third property of the normal use of language, namely its coherence and its "appropriateness to the situation" – which of course is an entirely different matter from control by external stimuli. Just what "appropriateness" and "coherence" may consist in we cannot say in any clear or definite way, but there is no doubt that these are meaningful concepts. We can distinguish normal use of language from the ravings of a maniac or the output of a computer with a random element.

Honesty forces us to admit that we are as far today as Descartes was three centuries ago from understanding just what enables a human to speak in a way that is innovative, free from stimulus control, and also appropriate and coherent. This is a serious problem that the psychologist and biologist must ultimately face and that cannot be talked out of existence by invoking "habit" or "conditioning" or "natural selection."

The Cartesian analysis of the problem of other minds, in terms of the creative aspect of language use and similar indications of the limits of mechanical explanation, was not entirely satisfying to contemporary opinion – Bayle's *Dictionary*, for example, cites the inability to give a satisfactory proof of the existence of other minds as the weakest element in the Cartesian philosophy – and there was a long and intriguing series of discussions and polemics regarding the problems that Descartes raised. From the vantage point of several centuries, we can see that the debate was inconclusive. The properties of human thought and human language emphasized by the Cartesians are real enough; they were then, as they are now, beyond the bounds of any well-understood kind of

physical explanation. Neither physics nor biology nor psychology gives us any clue as to how to deal with these matters.

As in the case of other intractable problems, it is tempting to try another approach, one that might show the problem to be misconceived, the result of some conceptual confusion. This is a line of argument that has been followed in contemporary philosophy, but, it seems to me, without success. It is clear that the Cartesians understood, as well as Gilbert Ryle and other contemporary critics understand, the difference between providing criteria for intelligent behavior, on the one hand, and providing an explanation for the possibility of such behavior, on the other; but, as distinct from Ryle, they were interested in the latter problem as well as the former. As scientists, they were not satisfied with the formulation of experimental tests that would show the behavior of another organism to be creative, in the special sense just outlined; they were also troubled, and quite rightly so, by the fact that the abilities indicated by such tests and observational criteria transcended the capacities of corporeal bodies as they understood them, just as they are beyond the scope of physical explanation as we understand it today. There is surely nothing illegitimate in an attempt to go beyond elaboration of observational tests and collection of evidence to the construction of some theoretical explanation for what is observed, and this is just what was at stake in the Cartesian approach to the problem of mind. As La Forge and others insisted, it is necessary to go beyond what one can perceive or "imagine" (in the technical, classical sense of this term) if one hopes to understand the nature of "l'esprit de l'homme," just as Newton did – successfully – in trying to understand the nature of planetary motion. On the other hand, the proposals of the Cartesians were themselves of no real substance; the phenomena in question are not explained satisfactorily by attributing them to an "active principle" called "mind," the properties of which are not developed in any coherent or comprehensive way.

It seems to me that the most hopeful approach today is to describe the phenomena of language and of mental activity as accurately as possible, to try to develop an abstract theoretical apparatus that will as far as possible account for these phenomena and reveal the principles of their organization and functioning, without attempting, for the present, to relate the postulated mental structures and processes to any physiological mechanisms or to interpret mental function in terms of "physical causes." We can only leave open for the future the question of how these abstract structures and processes are realized or accounted for in some concrete terms, conceivably in terms that are not within the range of physical processes as presently understood – a conclusion that, if correct, should surprise no one.

This rationalist philosophy of language merged with various other independent developments in the seventeenth century, leading to the first really significant general theory of linguistic structure, namely the general point of view that

came to be known as "philosophical" or "universal" grammar. Unfortunately, philosophical grammar is very poorly known today. There are few technical or scholarly studies, and these few are apologetic and disparaging. References to philosophical grammar in modern treatises on language are so distorted as to be quite worthless. Even a scholar with such high standards as Leonard Bloomfield gives an account of philosophical grammar in his major work, *Language*, that bears almost no resemblance to the original and attributes to this tradition views diametrically opposed to those that were most typical of it. For example, Bloomfield and many others describe philosophical grammar as based on a Latin model, as prescriptive, as showing no interest in the sounds of speech, as given to a confusion of speech with writing. All these charges are false, and it is important to dispel these myths to make possible an objective evaluation of what was actually accomplished.

It is particularly ironic that philosophical grammar should be accused of a Latin bias. In fact, it is significant that the original works – the Port-Royal *Grammar* and *Logic*, in particular – were written in French, the point being that they formed part of the movement to replace Latin by the vernacular. The fact is that Latin was regarded as an artificial and distorted language, one positively injurious to the exercise of the plain thinking and common-sense discourse by which the Cartesians set such store. The practitioners of philosophical grammar used such linguistic materials as were available to them; it is noteworthy that some of the topics that were studied with the greatest care and persistence for well over a century involved points of grammar that do not even have an analogue in Latin. A striking example is the so-called rule of Vaugelas, which involves the relation between indefinite articles and relative clauses in French. For 150 years the rule of Vaugelas was the central issue debated in the controversy over the possibility of developing a "rational grammar," one which would go beyond description to achieve a rational explanation for phenomena.

No doubt it is a complete misunderstanding of the issue of rational explanation that leads to the charge of "prescriptivism" that is leveled, quite erroneously, against philosophical grammar. In fact, there is no issue of prescriptivism. It was well understood and frequently reiterated that the facts of usage are what they are, and that it is not the place of the grammarian to legislate. At stake was an entirely different matter, namely the problem of accounting for the facts of usage on the basis of explanatory hypotheses concerning the nature of language and, ultimately, the nature of human thought. Philosophical grammarians had little interest in the accumulation of data, except insofar as such data could be used as evidence bearing on deeper processes of great generality. The contrast, then, is not between descriptive and prescriptive grammar, but between description and explanation, between grammar as "natural history" and grammar as a kind of "natural philosophy" or, in modern terms, "natural science." A largely irrational objection to explanatory theories as such has made it difficult for

modern linguistics to appreciate what was actually at stake in these developments and has led to a confusion of philosophical grammar with the effort to teach better manners to a rising middle class.

The whole matter is not without interest. I mentioned earlier that there are striking similarities between the seventeenth-century climate of opinion and that of contemporary cognitive psychology and linguistics. One point of similarity has to do with precisely this matter of explanatory theory. Philosophical grammar, very much like current generative grammar, developed in self-conscious opposition to a descriptive tradition that interpreted the task of the grammarian to be merely that of recording and organizing the data of usage – a kind of natural history. It maintained – quite correctly, I believe – that such a restriction was debilitating and unnecessary and that, whatever justification it may have, it has nothing to do with the method of science – which is typically concerned with data not for itself but as evidence for deeper, hidden organizing principles, principles that cannot be detected "in the phenomena" nor derived from them by taxonomic data-processing operations, any more than the principles of celestial mechanics could have been developed in conformity with such strictures.

Contemporary scholarship is not in a position to give a definitive assessment of the achievements of philosophical grammar. The ground-work has not been laid for such an assessment, the original work is all but unknown in itself, and much of it is almost unobtainable. For example, I have been unable to locate a single copy, in the United States, of the only critical edition of the Port-Royal *Grammar*, produced over a century ago; and although the French original is now once again available,[3] the one English translation of this important work is apparently to be found only in the British Museum. It is a pity that this work should have been so totally disregarded, since what little is known about it is intriguing and quite illuminating.

This is not the place to attempt a preliminary assessment of this work or even to sketch its major outlines as they now appear, on the basis of present, quite inadequate knowledge. However, I do want to mention at least a few of the persistent themes. It seems that one of the innovations of the Port-Royal *Grammar* of 1660 – the work that initiated the tradition of philosophical grammar – was its recognition of the importance of the notion of the phrase as a grammatical unit. Earlier grammar had been largely a grammar of word classes and inflections. In the Cartesian theory of Port-Royal, a phrase corresponds to a complex idea and a sentence is subdivided into consecutive phrases, which are further subdivided into phrases, and so on, until the level of the word is reached. In this way we derive what might be called the "surface structure" of the sentence in question. To use what became a standard example, the sentence "Invisible God created the visible world" contains the subject "invisible God"

[3] Menston, England: Scolar Press Limited, 1967.

and the predicate "created the visible world," the latter contains the complex idea "the visible world" and the verb "created," and so on. But it is interesting that although the Port-Royal *Grammar* is apparently the first to rely in a fairly systematic way on analysis into surface structure, it also recognized the inadequacy of such analysis. According to the Port-Royal theory, surface structure corresponds only to sound – to the corporeal aspect of language; but when the signal is produced, with its surface structure, there takes place a corresponding mental analysis into what we may call the deep structure, a formal structure that relates directly not to the sound but to the meaning. In the example just given, "Invisible God created the visible world," the deep structure consists of a system of three propositions, "that God is invisible," "that he created the world," "that the world is visible." The propositions that interrelate to form the deep structure are not, of course, asserted when the sentence is used to make a statement; if I say that a wise man is honest, I am not asserting that men are wise or honest, even though in the Port-Royal theory the propositions "a man is wise" and "a man is honest" enter into the deep structure. Rather, these propositions enter into the complex ideas that are present to the mind, though rarely articulated in the signal, when the sentence is uttered.

The deep structure is related to the surface structure by certain mental operations – in modern terminology, by grammatical transformations. Each language can be regarded as a particular relation between sound and meaning. Following the Port-Royal theory to its logical conclusions, then, the grammar of a language must contain a system of rules that characterizes deep and surface structures and the transformational relation between them, and – if it is to accommodate the creative aspect of language use – that does so over an infinite domain of paired deep and surface structures. To use the terminology Wilhelm von Humboldt used in the 1830s, the speaker makes infinite use of finite means. His grammar must, then, contain a finite system of rules that generates infinitely many deep and surface structures, appropriately related. It must also contain rules that relate these abstract structures to certain representations of sound and meaning – representations that, presumably, are constituted of elements that belong to universal phonetics and universal semantics, respectively. In essence, this is the concept of grammatical structure as it is being developed and elaborated today. Its roots are clearly to be found in the classical tradition that I am now discussing, and the basic concepts were explored with some success in this period.

The theory of deep and surface structure seems straightforward enough, at least in rough outline. Nevertheless, it was rather different from anything that preceded it, and, somewhat more surprising, it disappeared almost without a trace as modern linguistics developed in the late nineteenth century. I want to say just a word about the relationship of the theory of deep and surface structure to earlier and later thinking about language.

There is a similarity, which I think can be highly misleading, between the theory of deep and surface structure and a much older tradition. The practitioners of philosophical grammar were very careful to stress this similarity in their detailed development of the theory and had no hesitation in expressing their debt to classical grammar as well as to such major figures of renaissance grammar as the Spanish scholar Sanctius. Sanctius, in particular, had developed a theory of ellipsis that had great influence on philosophical grammar. As I have already remarked, philosophical grammar is poorly understood today. But such antecedents as Sanctius have fallen into total oblivion. Furthermore, as in the case of all such work, there is a problem of determining not only what he said but also, more importantly, what he meant.

There is no doubt that in developing his concept of ellipsis as a fundamental property of language, Sanctius gave many linguistic examples that superficially are closely parallel to those that were used to develop the theory of deep and surface structure, both in classical philosophical grammar and in its far more explicit modern variants. It means, however, that the concept of ellipsis is intended by Sanctius merely as a device for the interpretation of texts. Thus, to determine the true meaning of an actual literary passage one must very often, according to Sanctius, regard it as an elliptical variant of a more elaborate paraphrase. But the Port-Royal theory and its later development, particularly at the hands of the encyclopedist Du Marsais, gave a rather different interpretation to ellipsis. The clear intent of philosophical grammar was to develop a psychological theory, not a technique of textual interpretation. The theory holds that the underlying deep structure, with its abstract organization of linguistic forms, is "present to the mind," as the signal, with its surface structure, is produced or perceived by the bodily organs. And the transformational operations relating deep and surface structure are actual mental operations, performed by the mind when a sentence is produced or understood. The distinction is fundamental. Under the latter interpretation, it follows that there must be, represented in the mind, a fixed system of generative principles that characterize and associate deep and surface structures in some definite way – a grammar, in other words, that is used in some fashion as discourse is produced or interpreted. This grammar represents the underlying linguistic competence to which I referred earlier. The problem of determining the character of such grammars and the principles that govern them is a typical problem of science, perhaps very difficult, but in principle admitting of definite answers that are right or wrong as they do or do not correspond to the mental reality. But the theory of ellipsis as a technique of textual interpretation need not consist of a set of principles represented somehow in the mind as an aspect of normal human competence and intelligence. Rather, it can be in part ad hoc and can involve many cultural and personal factors relevant to the literary work under analysis.

The Port-Royal theory of deep and surface structure belongs to psychology as an attempt to elaborate Huarte's second type of wit, as an exploration of the properties of normal human intelligence. The concept of ellipsis in Sanctius, if I understand it correctly, is one of many techniques, to be applied as conditions warrant and having no necessary mental representation as an aspect of a normal intelligence. Although the linguistic examples used are often similar, the context in which they are introduced and the framework in which they fit are fundamentally different; in particular, they are separated by the Cartesian revolution. I propose this with some diffidence, because of the obscurity of the relevant texts and their intellectual backgrounds, but this interpretation seems to me correct.

The relation of the Port-Royal theory to modern structural and descriptive linguistics is somewhat clearer. The latter restricts itself to the analysis of what I have called surface structure, to formal properties that are explicit in the signal and to phrases and units that can be determined from the signal by techniques of segmentation and classification. This restriction is a perfectly self-conscious one, and it was regarded – I believe quite erroneously – as a great advance. The great Swiss linguist Ferdinand de Saussure, who at the turn of the century laid the groundwork for modern structural linguistics, put forth the view that the only proper methods of linguistic analysis are segmentation and classification. Applying these methods, the linguist determines the patterns into which the units so analyzed fall, where these patterns are either syntagmatic – that is, patterns of literal succession in the stream of speech – or paradigmatic – that is, relations among units that occupy the same position in the stream of speech. He held that when all such analysis is complete, the structure of the language is, of necessity, completely revealed, and the science of linguistics will have realized its task completely. Evidently, such taxonomic analysis leaves no place for deep structure in the sense of philosophical grammar. For example, the system of three propositions underlying the sentence "Invisible God created the visible world" cannot be derived from this sentence by segmentation and classification of segmented units, nor can the transformational operations relating the deep and surface structure, in this case, be expressed in terms of paradigmatic and syntagmatic structures. Modern structural linguistics has been faithful to these limitations, which were held to be necessary limitations.

In fact, Saussure in some respects even went beyond this in departing from the tradition of philosophical grammar. He occasionally expressed the view that processes of sentence formation do not belong to the system of language at all – that the system of language is restricted to such linguistic units as sounds and words and perhaps a few fixed phrases and a small number of very general patterns; the mechanisms of sentence formation are otherwise free from any constraint imposed by linguistic structure as such. Thus, in his terms, sentence

formation is not strictly a matter of *langue*, but is rather assigned to what he called *parole*, and thus placed outside the scope of linguistics proper; it is a process of free creation, unconstrained by linguistic rule except insofar as such rules govern the forms of words and the patterns of sounds. Syntax, in this view, is a rather trivial matter. And, in fact, there is very little work in syntax throughout the period of structural linguistics.

In taking this position, Saussure echoed an important critique of Humboldtian linguistic theory by the distinguished American linguist William Dwight Whitney, who evidently greatly influenced Saussure. According to Whitney, Humboldtian linguistic theory, which in many ways extended the Cartesian views that I have been discussing, was fundamentally in error. Rather, a language is simply "made up of a vast number of items, each of which has its own time, occasion, and effect." He maintained that "language in the concrete sense . . . is . . . the sum of words and phrases by which any man expresses his thought"; the task of the linguist, then, is to list these linguistic forms and to study their individual histories. In contrast to philosophical grammar, Whitney argued that there is nothing universal about the form of language and that one can learn nothing about the general properties of human intelligence from the study of the arbitrary agglomeration of forms that constitutes a human language. As he put it, "The infinite diversity of human speech ought alone to be a sufficient bar to the assertion that an understanding of the powers of the soul involves the explanation of speech." Similarly, Delbrück, in the standard work on Indo-European comparative syntax, denounced traditional grammar for having set up ideal sentence types underlying the observed signals, referring to Sanctius as the "major dogmatist in this domain."

With the expression of such sentiments as these, we enter the modern age of the study of language. The death-knell of philosophical grammar was sounded with the remarkable successes of comparative Indo-European studies, which surely rank among the outstanding achievements of nineteenth-century science. The impoverished and thoroughly inadequate conception of language expressed by Whitney and Saussure and numerous others proved to be entirely appropriate to the current stage of linguistic research. As a result, this conception was held to be vindicated, a not unnatural but thoroughly mistaken conviction. Modern structural-descriptive linguistics developed within the same intellectual framework and also made substantial progress, to which I will return directly. In contrast, philosophical grammar did not provide appropriate concepts for the new comparative grammar or for the study of exotic languages unknown to the investigator, and it was, in a sense, exhausted. It had reached the limits of what could be achieved within the framework of the ideas and techniques that were available. There was no clear understanding a century ago as to how one might proceed to construct generative grammars that "make infinite use of finite means" and that express the "organic form" of human language, "that

marvellous invention" (in the words of the Port-Royal *Grammar*) "by which we construct from twenty-five or thirty sounds an infinity of expressions, which, having no resemblance in themselves to what takes place in our minds, still enable us to let others know the secret of what we conceive and of all the various mental activities that we carry out."

Thus, the study of language had arrived at a situation in which there was, on the one hand, a set of simple concepts that provided the basis for some startling successes and, on the other, some deep but rather vague ideas that did not seem to lead to any further productive research. The outcome was inevitable and not at all to be deplored. There developed a professionalization of the field, a shift of interest away from the classical problems of general interest to intellectuals like Arnauld and Humboldt, for example, toward a new domain largely defined by the techniques that the profession itself has forged in the solution of certain problems. Such a development is natural and quite proper, but not without its dangers. Without wishing to exalt the cult of gentlemanly amateurism, one must nevertheless recognize that the classical issues have a liveliness and significance that may be lacking in an area of investigation that is determined by the applicability of certain tools and methods, rather than by problems that are of intrinsic interest in themselves.

The moral is not to abandon useful tools; rather, it is, first, that one should maintain enough perspective to be able to detect the arrival of that inevitable day when the research that can be conducted with these tools is no longer important; and, second, that one should value ideas and insights that are to the point, though perhaps premature and vague and not productive of research at a particular stage of technique and understanding. With the benefits of hindsight, I think we can now see clearly that the disparagement and neglect of a rich tradition proved in the long run to be quite harmful to the study of language. Furthermore, this disparagement and neglect were surely unnecessary. Perhaps it would have been psychologically difficult, but there is no reason in principle why the successful exploitation of the structuralist approach in historical and descriptive study could not have been coupled with a clear recognition of its essential limitations and its ultimate inadequacy, in comparison with the tradition it temporarily, and quite justifiably, displaced. Here, I think, lies a lesson that may be valuable for the future study of language and mind.

To conclude, I think there have been two really productive traditions of research that have unquestionable relevance to anyone concerned with the study of language today. One is the tradition of philosophical grammar that flourished from the seventeenth century through romanticism; the second is the tradition that I have rather misleadingly been referring to as "structuralist," which has dominated research for the past century, at least until the early 1950s. I have dwelt on the achievements of the former because of their unfamiliarity as well as their contemporary relevance. Structural linguistics has enormously broadened

the scope of information available to us and has extended immeasurably the reliability of such data. It has shown that there are structural relations in language that can be studied abstractly. It has raised the precision of discourse about language to entirely new levels. But I think that its major contribution may prove to be one for which, paradoxically, it has been very severely criticized. I refer to the careful and serious attempt to construct "discovery procedures," those techniques of segmentation and classification to which Saussure referred. This attempt was a failure – I think that is now generally understood. It was a failure because such techniques are at best limited to the phenomena of surface structure and cannot, therefore, reveal the mechanisms that underlie the creative aspect of language use and the expression of semantic content. But what remains of fundamental importance is that this attempt was directed to the basic question in the study of language, which was for the first time formulated in a clear and intelligible way. The problem raised is that of specifying the mechanisms that operate on the data of sense and produce knowledge of language – linguistic competence. It is obvious that such mechanisms exist. Children do learn a first language; the language that they learn is, in the traditional sense, an "instituted language," not an innately specified system. The answer that was proposed in structural linguistic methodology has been shown to be incorrect, but this is of small importance when compared with the fact that the problem itself has now received a clear formulation.

Whitehead once described the mentality of modern science as having been forged through "the union of passionate interest in the detailed facts with equal devotion to abstract generalization." It is roughly accurate to describe modern linguistics as passionately interested in detailed fact, and philosophical grammar as equally devoted to abstract generalization. It seems to me that the time has arrived to unite these two major currents and to develop a synthesis that will draw from their respective achievements. In the next two lectures, I will try to illustrate how the tradition of philosophical grammar can be reconstituted and turned to new and challenging problems and how one can, finally, return in a productive way to the basic questions and concerns that gave rise to this tradition.

2 Linguistic contributions to the study of mind: present

One difficulty in the psychological sciences lies in the familiarity of the phenomena with which they deal. A certain intellectual effort is required to see how such phenomena can pose serious problems or call for intricate explanatory theories. One is inclined to take them for granted as necessary or somehow "natural."

The effects of this familiarity of phenomena have often been discussed. Wolfgang Köhler, for example, has suggested that psychologists do not open up "entirely new territories" in the manner of the natural sciences, "simply because man was acquainted with practically all territories of mental life a long time before the founding of scientific psychology . . . because at the very beginning of their work there were no entirely unknown mental facts left which they could have discovered."[1] The most elementary discoveries of classical physics have a certain shock value – man has no intuition about elliptical orbits or the gravitational constant. But "mental facts" of even a much deeper sort cannot be "discovered" by the psychologist, because they are a matter of intuitive acquaintance and, once pointed out, are obvious.

There is also a more subtle effect. Phenomena can be so familiar that we really do not see them at all, a matter that has been much discussed by literary theorists and philosophers. For example, Viktor Shklovskij in the early 1920s developed the idea that the function of poetic art is that of "making strange" the object depicted. "People living at the seashore grow so accustomed to the murmur of the waves that they never hear it. By the same token, we scarcely ever hear the words which we utter . . . We look at each other, but we do not see each other any more. Our perception of the world has withered away; what has remained is mere recognition." Thus, the goal of the artist is to transfer what is depicted to the "sphere of new perception"; as an example, Shklovskij cites a story by Tolstoy in which social customs and institutions are "made strange" by the device of presenting them from the viewpoint of a narrator who happens to be a horse.[2]

[1] W. Köhler, *Dynamics in Psychology* (New York: Liveright, 1940).
[2] See V. Ehrlich, *Russian Formalism*, 2nd rev. edn. (New York: Humanities, 1965), pp. 176–77.

The observation that "we look at each other, but we do not see each other any more" has perhaps itself achieved the status of "words which we utter but scarcely ever hear." But familiarity, in this case as well, should not obscure the importance of the insight.

Wittgenstein makes a similar observation, pointing out that "the aspects of things that are most important for us are hidden because of their simplicity and familiarity (one is unable to notice something – because it is always before one's eyes)."[3] He sets himself to "supplying . . . remarks on the natural history of human beings: we are not contributing curiosities however, but observations which no one has doubted, but which have escaped remark only because they are always before our eyes."[4]

Less noticed is the fact that we also lose sight of the need for explanation when phenomena are too familiar and "obvious." We tend too easily to assume that explanations must be transparent and close to the surface. The greatest defect of classical philosophy of mind, both rationalist and empiricist, seems to me to be its unquestioned assumption that the properties and content of the mind are accessible to introspection; it is surprising to see how rarely this assumption has been challenged, insofar as the organization and function of the intellectual faculties are concerned, even with the Freudian revolution. Correspondingly, the far-reaching studies of language that were carried out under the influence of Cartesian rationalism suffered from a failure to appreciate either the abstractness of those structures that are "present to the mind" when an utterance is produced or understood, or the length and complexity of the chain of operations that relate the mental structures expressing the semantic content of the utterance to the physical realization.

A similar defect mars the study of language and mind in the modern period. It seems to me that the essential weakness in the structuralist and behaviorist approaches to these topics is the faith in the shallowness of explanations, the belief that the mind must be simpler in its structure than any known physical organ and that the most primitive of assumptions must be adequate to explain whatever phenomena can be observed. Thus, it is taken for granted without argument or evidence (or is presented as true by definition) that a language is a "habit structure" or a network of associative connections, or that knowledge of language is merely a matter of "knowing how," a skill expressible as a system of dispositions to respond. Accordingly, knowledge of language must develop slowly through repetition and training, its apparent complexity resulting from the proliferation of very simple elements rather than from deeper principles of mental organization that may be as inaccessible to introspection

[3] Ludwig Wittgenstein, *Philosophical Investigations* (New York: Oxford University Press, 1953), Section 129.
[4] *Ibid.*, Section 415.

as the mechanisms of digestion or coordinated movement. Although there is nothing inherently unreasonable in an attempt to account for knowledge and use of language in these terms, it also has no particular plausibility or a priori justification. There is no reason to react with uneasiness or disbelief if study of the knowledge of language and use of this knowledge should lead in an entirely different direction.

I think that in order to achieve progress in the study of language and human cognitive faculties in general it is necessary first to establish "psychic distance" from the "mental facts" to which Köhler referred, and then to explore the possibilities for developing explanatory theories, whatever they may suggest with regard to the complexity and abstractness of the underlying mechanisms. We must recognize that even the most familiar phenomena require explanation and that we have no privileged access to the underlying mechanisms, no more so than in physiology or physics. Only the most preliminary and tentative hypotheses can be offered concerning the nature of language, its use, and its acquisition. As native speakers, we have a vast amount of data available to us. For just this reason it is easy to fall into the trap of believing that there is nothing to be explained, that whatever organizing principles and underlying mechanisms may exist must be "given" as the data is given. Nothing could be further from the truth, and an attempt to characterize precisely the system of rules we have mastered that enables us to understand new sentences and produce a new sentence on an appropriate occasion will quickly dispel any dogmatism on this matter. The search for explanatory theories must begin with an attempt to determine these systems of rules and to reveal the principles that govern them.

The person who has acquired knowledge of a language has internalized a system of rules that relate sound and meaning in a particular way. The linguist constructing a grammar of a language is in effect proposing a hypothesis concerning this internalized system. The linguist's hypothesis, if presented with sufficient explicitness and precision, will have certain empirical consequences with regard to the form of utterances and their interpretations by the native speaker. Evidently, knowledge of language – the internalized system of rules – is only one of the many factors that determine how an utterance will be used or understood in a particular situation. The linguist who is trying to determine what constitutes knowledge of a language – to construct a correct grammar – is studying one fundamental factor that is involved in performance, but not the only one. This idealization must be kept in mind when one is considering the problem of confirmation of grammars on the basis of empirical evidence. There is no reason why one should not also study the interaction of several factors involved in complex mental acts and underlying actual performance, but such a study is not likely to proceed very far unless the separate factors are themselves fairly well understood.

In a good sense, the grammar proposed by the linguist is an explanatory theory; it suggests an explanation for the fact that (under the idealization mentioned) a speaker of the language in question will perceive, interpret, form, or use an utterance in certain ways and not in other ways. One can also search for explanatory theories of a deeper sort. The native speaker has acquired a grammar on the basis of very restricted and degenerate evidence; the grammar has empirical consequences that extend far beyond the evidence. At one level, the phenomena with which the grammar deals are explained by the rules of the grammar itself and the interaction of these rules. At a deeper level, these same phenomena are explained by the principles that determine the selection of the grammar on the basis of the restricted and degenerate evidence available to the person who has acquired knowledge of the language, who has constructed for himself this particular grammar. The principles that determine the form of grammar and that select a grammar of the appropriate form on the basis of certain data constitute a subject that might, following a traditional usage, be termed "universal grammar." The study of universal grammar, so understood, is a study of the nature of human intellectual capacities. It tries to formulate the necessary and sufficient conditions that a system must meet to qualify as a potential human language, conditions that are not accidentally true of the existing human languages, but that are rather rooted in the human "language capacity," and thus constitute the innate organization that determines what counts as linguistic experience and what knowledge of language arises on the basis of this experience. Universal grammar, then, constitutes an explanatory theory of a much deeper sort than particular grammar, although the particular grammar of a language can also be regarded as an explanatory theory.[5]

In practice, the linguist is always involved in the study of both universal and particular grammar. When he constructs a descriptive, particular grammar in one way rather than another on the basis of what evidence he has available, he is guided, consciously or not, by certain assumptions as to the form of grammar, and these assumptions belong to the theory of universal grammar. Conversely, his formulation of principles of universal grammar must be justified by the study of their consequences when applied in particular grammars. Thus, at several levels the linguist is involved in the construction of explanatory theories, and at each level there is a clear psychological interpretation for his theoretical and descriptive work. At the level of particular grammar, he is attempting to characterize knowledge of a language, a certain cognitive system that has been developed – unconsciously, of course – by the normal speaker–hearer. At the level of universal grammar, he is trying to establish certain general properties

[5] To bring out this difference in depth of explanation, I have suggested in my *Current Issues in Linguistic Theory* (New York: Humanities, 1965) that the term "level of descriptive adequacy" might be used for the study of the relation between grammars and data and the term "level of explanatory adequacy" for the relation between a theory of universal grammar and these data.

of human intelligence. Linguistics, so characterized, is simply the subfield of psychology that deals with these aspects of mind.

I will try to give some indication of the kind of work now in progress that aims, on the one hand, to determine the systems of rules that constitute knowledge of a language, and on the other, to reveal the principles that govern these systems. Obviously, any conclusions that can be reached today regarding particular or universal grammar must be quite tentative and restricted in their coverage. And in a brief sketch such as this only the roughest outlines can be indicated. To try to give something of the flavor of what is being done today I will concentrate on problems that are current in that they can be formulated with some clarity and studied, though they still resist solution.

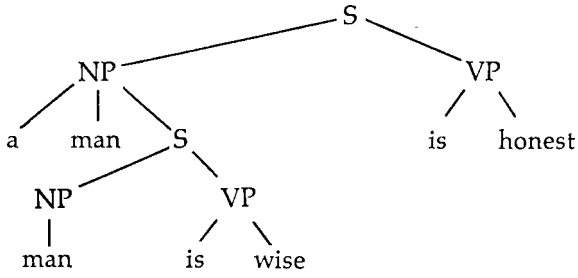
As I indicated in the first lecture, I believe that the most appropriate general framework for the study of problems of language and mind is the system of ideas developed as part of the rationalist psychology of the seventeenth and eighteenth centuries, elaborated in important respects by the romantics and then largely forgotten as attention shifted to other matters. According to this traditional conception, a system of propositions expressing the meaning of a sentence is produced in the mind as the sentence is realized as a physical signal, the two being related by certain formal operations that, in current terminology, we may call *grammatical transformations.* Continuing with current terminology, we can thus distinguish the *surface structure* of the sentence, the organization into categories and phrases that is directly associated with the physical signal, from the underlying *deep structure*, also a system of categories and phrases, but with a more abstract character. Thus, the surface structure of the sentence "A wise man is honest" might analyze it into the subject "a wise man" and the predicate "is honest." The deep structure, however, will be rather different. It will, in particular, extract from the complex idea that constitutes the subject of the surface structure an underlying proposition with the subject "man" and the predicate "be wise." In fact, the deep structure, in the traditional view, is a system of two propositions, neither of which is asserted, but which interrelate in such a way as to express the meaning of the sentence "A wise man is honest." We might represent the deep structure in this sample case by formula 1, and the surface structure by formula 2, where paired brackets are labeled to show the category of phrase that they bound. (Many details are omitted.)

1 $$\left[_{S}\left[_{NP}\left[\begin{array}{l}\text{a man}\\ S\end{array}\left[_{NP}[\text{man}]_{NP}\ VP[\text{is wise}]_{VP}\right]_{S}\right]_{NP}\ VP[\text{is honest}]_{VP}\right]_{S}\right]$$

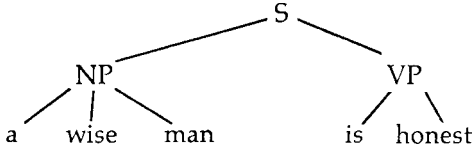
2 $$\left[_{S}\left[_{NP}[\text{a wise man}]_{NP}\ VP\ [\text{is honest}]_{VP}\right]_{S}\right]$$

An alternative and equivalent notation, widely used, expresses the labeled bracketing of 1 and 2 in tree form, as 1′ and 2′ respectively:

1′



2′



If we understand the relation "subject-of" to hold between a phrase of the category noun phrase (NP) and the sentence (S) that directly dominates it, and the relation "predicate-of" to hold between a phrase of the category verb phrase (VP) and the sentence that directly dominates it, then structures 1 and 2 (equivalently, 1′ and 2′) specify the grammatical functions of subject and predicate in the intended way. The grammatical functions of the deep structure (1) play a central role in determining the meaning of the sentence. The phrase structure indicated in 2, on the other hand, is closely related to its phonetic shape – specifically, it determines the intonation contour of the utterance represented.

Knowledge of a language involves the ability to assign deep and surface structures to an infinite range of sentences, to relate these structures appropriately, and to assign a semantic interpretation and a phonetic interpretation to the paired deep and surface structures. This outline of the nature of grammar seems to be quite accurate as a first approximation to the characterization of "knowledge of a language."

How are the deep and surface structures related? Clearly, in the simple example given, we can form the surface structure from the deep structure by performing such operations as the following:

3
a. assign the marker *wh-* to the most deeply embedded NP, "man"
b. replace the NP so marked by "who"
c. delete "who is"
d. invert "man" and "wise."

Applying just operations a and b, we derive the structure underlying the sentence "a man who is wise is honest," which is one possible realization of the

underlying structure (1). If, furthermore, we apply the operation c (deriving "a man wise is honest"), we must, in English, also apply the subsidiary operation d, deriving the surface structure (2), which can then be phonetically interpreted.

If this approach is correct in general, then a person who knows a specific language has control of a grammar that *generates* (that is, characterizes) the infinite set of potential deep structures, maps them onto associated surface structures, and determines the semantic and phonetic interpretations of these abstract objects.[6] From the information now available, it seems accurate to propose that the surface structure determines the phonetic interpretation completely and that the deep structure expresses those grammatical functions that play a role in determining the semantic interpretation, although certain aspects of the surface structure may also participate in determining the meaning of the sentence in ways that I will not discuss here. A grammar of this sort will therefore define a certain infinite correlation of sound and meaning. It constitutes a first step toward explaining how a person can understand an arbitrary sentence of his language.

Even this artificially simple example serves to illustrate some properties of grammars that appear to be general. An infinite class of deep structures much like 1 can be generated by very simple rules that express a few rudimentary grammatical functions, if we assign to these rules a recursive property – in particular, one that allows them to embed structures of the form [s . . .]s within other structures. Grammatical transformations will then iterate to form, ultimately, a surface structure that may be quite remote from the underlying deep structure. The deep structure may be highly abstract; it may have no close point-by-point correlation to the phonetic realization. Knowledge of a language – "linguistic competence," in the technical sense of this term discussed briefly in the first lecture – involves a mastery of these grammatical processes.

With just this much of a framework, we can begin to formulate some of the problems that call for analysis and explanation. One major problem is posed by the fact that the surface structure generally gives very little indication in itself of the meaning of the sentence. There are, for example, numerous sentences

[6] For a detailed development of this point of view, see J. Katz and P. Postal, *An Integrated Theory of Linguistic Descriptions* (Cambridge, Mass.: MIT Press, 1964) and my *Aspects of the Theory of Syntax* (Cambridge, Mass.: MIT Press, 1965). See also Peter S. Rosenbaum, *The Grammar of English Predicate Complement Constructions* (Cambridge, Mass.: MIT Press, 1967). These contain references to earlier work that they extend and modify. There has been a great deal of work in the past few years extending and modifying this general approach still further and exploring alternatives. At present the field is in considerable ferment, and it will probably be some time before the dust begins to settle and a number of outstanding issues are even tentatively resolved. Current work is too extensive for detailed reference to be called for in a sketch such as this. Some idea of its scope and general directions can be obtained from collections such as R. Jacobs and P. S. Rosenbaum, eds., *Readings in English Transformational Grammar* (Waltham, Mass.: Ginn and Company, 1970).

that are ambiguous in some way that is not indicated by the surface structure. Consider sentence 4:

4 I disapprove of John's drinking.

This sentence can refer either to the fact of John's drinking or to its character. The ambiguity is resolved, in different ways, in sentences 5 and 6:

5 I disapprove of John's drinking the beer.

6 I disapprove of John's excessive drinking.

It is clear that grammatical processes are involved. Notice that we cannot simultaneously extend 4 in both of the ways illustrated in 5 and 6; that would give us 7:

7 *I disapprove of John's excessive drinking the beer.[7]

Our internalized grammar assigns two different abstract structures to 4, one of which is related to the structure that underlies 5, the other to the structure that underlies 6. Bu it is at the level of deep structure that the distinction is represented; it is obliterated by the transformations that map the deep structures onto the surface form associated with 4.

The processes that are involved in examples 4, 5, and 6 are quite common in English. Thus, the sentence "I disapprove of John's cooking" may imply either that I think his wife should cook or that I think he uses too much garlic, for example. Again, the ambiguity is resolved if we extend the sentence in the manner indicated in 5 and 6.

The fact that 7 is deviant requires explanation. The explanation in this case would be provided, at the level of particular grammar, by formulation of the grammatical rules that assign alternative deep structures and that in each case permit one but not the other of the extensions to 5 or 6. We would then explain the deviance of 7 and the ambiguity of 4 by attributing this system of rules to the person who knows the language, as one aspect of his knowledge. We might, of course, try to move to a deeper level of explanation, asking how it is that the person has internalized these rules instead of others that would determine a different sound–meaning correlation and a different class of generated surface structures (including, perhaps, 7). This is a problem of universal grammar, in the sense described earlier. Using the terminology of note 5, the discussion at the level of particular grammar would be one of descriptive adequacy, and at the level of universal grammar it would be one of explanatory adequacy.

Notice that the internalized rules of English grammar have still further consequences in a case like the one just discussed. There are transformations of great generality that permit or require the deletion of repeated elements, in whole

[7] I use the asterisk in the conventional way, to indicate a sentence that deviates in some respect from grammatical rule.

or in part, under well-defined conditions. Applied to structure 8, these rules derive 9.[8]

8 I don't like John's cooking any more than Bill's cooking.

9 I don't like John's cooking any more than Bill's.

Sentence 9 is ambiguous. It can mean either that I don't like the fact that John cooks any more than I like the fact that Bill cooks, or that I don't like the quality of John's cooking any more than I like the quality of Bill's cooking.[9] However, it cannot mean that I don't like the quality of John's cooking any more than I like the fact that Bill cooks, or conversely, with "fact" and "quality" interchanged. That is, in the underlying structure (8) we must understand the ambiguous phrases "John's cooking" and "Bill's cooking" in the same way if we are to be able to delete "cooking." It seems reasonable to assume that what is involved is some general condition on the applicability of deletion operations such as the one that gives 9 from 8, a rather abstract condition that takes into account not only the structure to which the operation applies but also the history of derivation of this structure.

Other examples can be found where a similar principle seems to be at work. Thus, consider sentence 10, which is presumably derived from either 11 or 12 and is therefore ambiguous:[10]

10 I know a taller man than Bill.

11 I know a taller man than Bill does.

12 I know a taller man than Bill is.

[8] Henceforth I shall generally delete brackets in giving a deep, surface, or intermediate structure, where this will not lead to confusion. One should think of 8 and 9 as each having a full labeled bracketing associated with it. Notice that 8 is not, of course, a deep structure, but rather the result of applying transformations to a more primitive abstract object.

[9] There may also be other interpretations, based on other ambiguities in the structure "John's cooking" – specifically, the cannibalistic interpretation and the interpretation of "cooking" as "that which is cooked."

[10] I should emphasize that when I speak of a sentence as derived by transformation from another sentence, I am speaking loosely and inaccurately. What I should say is that the structure associated with the first sentence is derived from the structure underlying the second. Thus, in the case now being discussed, it is the surface structure of 10 that is derived, on one analysis, from the abstract structure which, were it to undergo a different transformational development, would be converted into the surface structure of 11. That sentences are not derived from other sentences but rather from the structures underlying them has been explicitly assumed since the earliest work in transformational generative grammar about fifteen years ago, but informal statements such as those in the text here have misled many readers and have led to a good deal of confusion in the literature. Adding to the confusion, perhaps, is the fact that a very different theory of transformational relations developed by Zellig Harris, Henry Hiz, and others, does in fact regard the transformational operations as applied to sentences. See, for example, Z. S. Harris, "Co-occurrence and Transformation in Linguistic Structure," in *Language*, Vol. 33, No. 3, 1957, pp. 283-340, and many later publications. For me, and most other speakers, sentence 12 is deviant. Nevertheless, the association structure that underlies 10 under one analysis must be postulated, perhaps deriving from the structure associated with "I know a man who is taller than Bill is."

It seems clear that the ambiguity of 10 is not represented in the surface structure; the deletion of "does" in 11 leaves exactly the same structure as the deletion of "is" in 12. But now consider sentence 13.

13 I know a taller man than Bill, and so does John.

This sentence, like 9, is two-ways ambiguous rather than four-ways ambiguous. It can have the meaning of either 14 or 15, but not 16 or 17:[11]

14 I know a taller man than Bill does and John knows a taller man than Bill does.

15 I know a taller man than Bill is and John knows a taller man than Bill is.

16 I know a taller man than Bill is and John knows a taller man than Bill does.

17 I know a taller man than Bill does and John knows a taller man than Bill is.

But now a problem arises, as we can see by considering more carefully the derivation of 13. Let us refer to the deletion operation that gives 10 from 11 as T_1, and to the deletion operation that gives 10 from 12 as T_2. If we apply T_1 to each of the conjuncts of 14, we derive 18:

18 I know a taller man than Bill and John knows a taller man than Bill.

Application of T_2 to each of the conjuncts of 15 will also yield 18. But application of T_1 to one conjunct and T_2 to the other conjunct in 16 will also give 18, as will the same procedure (in the opposite order) when applied to the two conjuncts of 17. Thus, 18 can be derived by application of T_1 and T_2 to any of the four underlying forms, 14, 15, 16, or 17. The structure of 18 itself does not indicate which of these is the underlying form; the distinction has been eliminated by the deletion operations T_1 and T_2. But now consider the operation T_3, which derives "I saw Bill and so did John" from "I saw Bill and John saw Bill." Applying T_3 to 18, we derive 13. However, we have noted that 13 can

[11] It also cannot have the meaning "I know a taller man than Bill and John likes ice cream." Hence, if deep structure determines meaning (insofar as grammatical relations are involved), it must be that something like 14 or 15 is the immediately underlying structure for 13. It is a general property of deletion operations that some sort of recoverability is involved, a nontrivial matter with interesting empirical consequences. For some discussion, see my *Current Issues*, Section 2.2, and *Aspects*, Section 4.2.2. The problem posed by such examples as 9 and 13 was pointed out to me by John Ross. The first reference to the possibility that history of derivation may play a role in determining applicability of transformations appears in R. B. Lees, *The Grammar of English Nominalizations* (New York: Humanities, 1960), p. 76, in connection with his discussion – also the first – of the problem of identity of constituent structure as a factor in determining applicability of transformations.

have the interpretation of 14 or 15, but not of 16 or 17. Thus we can see that T_3 can apply to 18 only if either 14 or 15, but not 16 or 17, was the structure underlying 18 in the given derivations of 18. However, this information is not represented in 18 itself, as we have just observed. Therefore, to apply T_3 to 18 we must know something about the history of derivation of 18 – we must have information that would not be contained in the labeled bracketing of 18 itself. What we must know, in fact, is that the two conjuncts of 18 derive from underlying structures in which the same element was deleted.[12] It appears, once again, that some general condition on applicability of deletion transformations must be involved, a principle that somehow brings into consideration the history of derivation of deleted strings, perhaps certain properties of the deep structure from which they ultimately derive.

To see how complex the problem is, consider such sentences as "John's intelligence, which is his most remarkable quality, exceeds his prudence" or "The book, which weighs five pounds, was written by John." Presumably, the relative pronoun in the embedded appositive clause replaces a deleted noun phrase, and the condition on deletion that we are discussing implies that this noun phrase should be identical to the antecedent noun phrase "John's intelligence" or "the book" in the underlying structure of the appositive clause. In each case, however, it can be argued that there is a difference between the antecedent and the noun phrase of the appositive clause. Thus, in the first case, we are referring to the degree of John's intelligence in the main clause but to the quality of his intelligence in the embedded clause; and in the second case we are referring to the book as an abstract object in the main clause but as a concrete physical object in the embedded clause; one might expect these differences to be represented in deep structure, thus contradicting the principle to which we seem to be driven by the earlier examples. I will not go on with this discussion here, but the reader will discover, if he pursues the matter, that the problem is compounded when a richer class of cases is considered.

In fact, the correct principle is unknown in such cases as these, although some of the conditions it must meet are clear. The problem posed by these examples is a quite typical one. Attention to linguistic fact reveals certain properties of sentences, relating to their sound, their meaning, their deviance, and so on. Evidently no explanation for these facts will be forthcoming so long as we restrict ourselves to vague talk about "habits" and "skills" and "dispositions to respond," or about the formation of sentences "by analogy." We do not have the "habit" of understanding sentences 4, 9, and 13 in a certain way: it is unlikely that the reader has ever encountered sentences closely resembling these, but he understands them in a highly specific way nevertheless. To refer to

[12] If 18 itself is only two-ways ambiguous, a problem in fact arises at an even earlier point. The unnaturalness of 18 makes it difficult to determine this with any confidence.

the processes involved as "analogy" is simply to give a name to what remains a mystery. To explain such phenomena we must discover the rules that relate sound and meaning in the language in question – the grammar that has been internalized by the person who knows the language – and the general principles that determine the organization and function of these rules.

The misleading and inadequate character of surface structure becomes evident as soon as even the most simple patterns are studied. Consider, for example, sentence 19 – again, an artificially simple example:

19 John was persuaded to leave.

The deep structure underlying this sentence must indicate that the subject–predicate relation holds in an underlying proposition of the form of 20 (assuming grammatical functions to be represented in the same manner suggested earlier), and that the verb–object relation holds in an underlying proposition of the form of 21:

20 $\left[_{S}\left[_{NP}{}^{John}\right]_{NP}\right]_{VP}{}^{leave}\left]_{VP}\right]_{S}$

21 $\left[_{S}\left[_{NP}\cdots\right]_{NP}\left[_{VP}{}^{persuade}\left[_{NP}{}^{John}\right]_{NP}\right]_{VP}\right]_{S}$

Thus, "John" is understood to be the subject of "leave" and the object of "persuade" in 19, and these facts are properly expressed in the deep structure underlying 19 if this deep structure embodies the propositions informally represented as 20 and 21. Although the deep structure must be constituted of such propositions, if the approach loosely outlined earlier is correct, there is no trace of them in the surface structure of the utterance. The various transformations that produce 19 have thoroughly obliterated the system of grammatical relations and functions that determine the meaning of the sentence.

The point becomes still more obvious if we take note of the variety of sentences that seem superficially to resemble 19, but that differ widely in the ways they are understood and the formal operations that apply to them. Suppose that "persuaded" in 19 is replaced by one of the following words:[13]

22 expected, hired, tired, pleased, happy, lucky, eager, certain, easy

With "expected" replacing "persuaded," the sentence can mean roughly that the fact of John's leaving was expected; but it is impossible to speak of the fact of John's leaving being persuaded. With "hired," the sentence has an entirely different meaning, roughly that the purpose of hiring John was so that he would leave – an interpretation that becomes more natural if we replace "leave" by a phrase like "fix the roof." When "tired" is substituted, we derive a nonsentence;

[13] See R. B. Lees, "A Multiply Ambiguous Adjectival Construction in English," in *Language*, Vol. 36, No. 2, 1960, pp. 207-21, for a discussion of such structures.

it becomes a sentence if "too tired" replaces "persuaded," the sentence now implying that John didn't leave. The word "pleased" is still different. In this case we can have "too pleased," implying that John didn't leave, but we can also extend the sentence to "John was too pleased to leave to suit me," which is impossible in the earlier cases. "Happy" is rather like "pleased," though one might argue that the verb–object relation holds between "please" and "John." The sentence "John was lucky to leave" is interpreted in still another way. It means, roughly, that John was lucky in that he left, an interpretation that is impossible in the earlier cases; furthermore, we can construct such sentences as "John was a lucky fellow to leave (so early)," but none of the earlier examples can replace "lucky" in such sentences. "John was eager to leave" differs from the earlier cases in that it is formally associated with such expressions as "John was eager for Bill to leave" and "John's eagerness (for Bill) to leave." "John was certain to leave" can be paraphrased as "it was certain that John would leave"; of the other examples, only "expected" is subject to this interpretation, but "expected" obviously differs from "certain" in numerous other respects – for example, it appears in a sentence such as "They expected John to leave." The word "easy" is of course entirely different; in this and only this case the verb–object relation holds between "leave" and "John."

It is clear, in short, that the surface structure is often misleading and uninformative and that our knowledge of language involves properties of a much more abstract nature, not indicated directly in the surface structure. Furthermore, even such artificially simple examples as these show how hopeless it would be to try to account for linguistic competence in terms of "habits," "dispositions," "knowing how," and other concepts associated with the study of behavior, as this study has been circumscribed, quite without warrant, in recent years.

Even at the level of sound structure, there is evidence that abstract representations are formed and manipulated in the mental operations involved in language use. We have a more detailed understanding of the nature of linguistic representation and the intricate conditions on rule application in this domain than in any other. The work of the past few years on sound structure seems to me to provide substantial evidence in support of the view that the form of particular grammars is determined, in highly significant ways, by a restrictive schematism that specifies the choice of relevant phonetic properties, the kinds of rules that can relate surface structure to phonetic representation, and the conditions on organization and application of these rules. It thus relates closely to the general topics discussed in the first lecture, topics that I will take up again below in considering the question of how this restrictive, universal schematism comes to be used in language acquisition. Furthermore, these investigations of sound structure, insofar as they support the conclusion that abstract phonological structures are manipulated by tightly organized and intricate systems of rules, are relevant to the very interesting problem of developing empirically

adequate models of performance. They suggest that all current approaches to problems of perception and organization of behavior suffer from a failure to attribute sufficient depth and complexity to the mental processes that must be represented in any model that attempts to come to grips with the empirical phenomena. Space does not permit a detailed development of these topics, either with respect to the matter of phonological structure or with respect to its potential significance for cognitive psychology.[14] However, one simple illustrative example, which is quite typical, may give some idea of the nature of the evidence that is available and the conclusions to which it points.

Recall that the syntactic rules of the language generate an infinite set of surface structures, each of which is a labeled bracketing of a string of minimal elements, such as 2, in which we may take the minimal elements to be the items *a*, *wise*, *man*, *is*, honest. Each of these items can itself be represented as a string of segments, for example *man* as the string of segments /m/, /æ/, /n/. Each of these segments may be regarded in turn as a set of specified features; thus, /m/ stands for the feature complex [+ consonantal], [– vocalic], [+ nasal], and so on. The segmental constitution of an item will be given by a lexical entry – a characterization of the inherent phonetic, semantic, and syntactic properties of the items in question. The lexicon of the language is the set of such lexical entries, with, perhaps, additional structure that need not concern us here. We are concerned now only with the phonetic properties of the lexical entry.

The lexical entry of an item must specify just those properties that are idiosyncratic, that are not determined by linguistic rule. For example, the lexical entry for *man* must indicate that its second segment is a low front vowel, but the degree of tenseness, diphthongization, nasalization, and so on, of this vowel need not be indicated in the lexical entry, since these are a matter of general rule, in part particular to various English dialects, in part common to all English dialects, in part a matter of universal phonology. Similarly, the lexical entry for *man* must indicate that it has an irregular plural, with the vowel shifting from low to mid. The segments of the lexical entry are abstract in the sense that the phonological rules of the language will frequently modify and elaborate them in a variety of ways; hence there need not be, in general, a simple point-by-point correspondence between the lexical entry and the actual phonetic representation. In discussing examples, I will use phonetic symbols in the usual way, each being regarded as a complex of a certain set of features. I will use the symbol / / to

[14] For discussion of these topics, see my article "Some General Properties of Phonological Rules," in *Language*, Vol. 47, No. 1, 1967. For a much fuller and more detailed discussion of phonological theory and its application to English, with examples drawn from many languages and some discussion of the history of the English sound system as well, see N. Chomsky and M. Halle, *The Sound Pattern of English* (New York: Harper & Row, 1968). The example in the text is discussed in detail, in the context of a more general framework of rules and principles, in Chapter 4, Section 4, of *The Sound Pattern of English.* See P. Postal, *Aspects of Phonological Theory* (New York: Harper & Row, 1968), for a general development of many related topics, along with a critical analysis of alternative approaches to the study of sound structure.

enclose lexical representations, and the symbol [ ] to enclose all representations derived from lexical representations by application of phonological rules, including, in particular, the final phonetic representation derived by application of the full set of phonological rules.

Consider first such words as *sign-signify*, *paradigm-paradigmatic*, and so on. For reasons that will become clearer as we proceed, it is the derived form, in this case, that is most closely related to the underlying abstract lexical representation. Suppose, then, that we tentatively assign to the stem in these forms the lexical representation /sign/ and /pærædigm/ where the symbols have their conventional phonetic interpretation. Thus, the underlying element /sign/ is realized as phonetic [sign] before *-ify.* However, it is realized as phonetic [sayn] in isolation. A similar observation holds of *paradigm.*

The forms of *sign* and *paradigm* in isolation are determined by certain phonological rules that, operating jointly, have the effect of converting the representation /ig/ to [ay] when followed by a word-final nasal. A careful analysis of English phonology shows that this process can be broken into a sequence of steps, including the following (the second and third of which, in fact, require further analysis).

23
a. velar becomes continuant before word-final nasal
b. vowel + velar continuant becomes tense vowel
c. /ī/ becomes [ay] (where /ī/ is the tense segment corresponding to [i])

Applying these rules to underlying /sign/ in isolation, we derive first [siγn] (where [γ] is the velar continuant) by 23a; then [sīn] by 23b; and finally [sayn] by 24c.

Rules 23a and 23b are of little interest, but 23c is a part of a very general system of rules of "vowel shift" that is quite central to English phonology. There are, for example, strong reasons for supposing that the stem underlying the forms *divine-divinity* is /divīn/, where the segment /ī/ is weakened to [i] before *-ity* and becomes [ay] by rule 23c in isolation. Similarly, *reptile* derives from underlying /reptīl/, which becomes [reptayl] by 23c in isolation and [reptil] before *-ian*, with the same shortening of vowel that takes place in *divinity*; and so on, in many other cases.

Consider next such words as *ignite-ignition*, *expedite-expeditious*, and *contrite-contrition.* Just as *reptile* and *divine* derive, by vowel shift, from /reptīl/ and /divīn/, so we can derive the first member of each of these pairs from /ignīt/, /expedīt/, and /contrīt/, respectively. The rule that applies to give the phonetic realization is 23c, a special case of the general process of vowel shift. Evidently, the second member of each pair is derived by such processes as 24 and 25:

24 Vowels become nontense before *-ion*, *-ious*, *-ian*, *-ity*, and so on.

25 The segment /t/ followed by a high front vowel is realized as [š].

The first of these rules is the one that gives [divin] from /divīn/ in *divinity* and [reptil] from /reptīl/ in *reptilian.* Similarly, it gives [ignit] from /ignīt/ in *ignition*, [expedit] from /expedīt/ in *expeditious*, and [contrit] from /contīt/ in *contrition.* There is an obvious underlying generalization, namely that a vowel becomes nontense before an unstressed vowel that is not in a word-final syllable; when properly formulated, this rule, along with vowel shift and a few others, constitutes the central portion of the English phonological system.

The second rule, 25, applies to the element /ti/ in /ignition/, /expeditious/, and /contrition/, replacing it by [š] and giving, finally, the phonetic realizations [ignišən], [ekspədišəs], [kəntrišən], after the application of the rule that reduces unstressed vowels to [ə]. In short, the segments realized as [ayt] in *ignite*, *expedite*, and *contrite* are realized as [iš] in *ignition*, *expeditious*, and *contrition.*

But now consider the words *right-righteous*, phonetically [rayt]-[rayčəs]. The latter form appears to deviate from the regular pattern in two respects, namely in vowel quality (we would expect [i] rather than [ay], by rule 24), and in the final consonant of the stem (we would expect [š] rather than [č], by rule 25). If *right* were subject to the same processes as *expedite*, we would have [rišəs] rather than [rayčəs] as the phonetic realization, analogous to [ekspədišəs]. What is the explanation for this double deviation?

Notice first that rule 25 is not quite exact; there are, in fact, other cases in which /ti/ is realized as [č] rather than as [š], for example *question* [kweschən], contrasted with *direction* [dərekšən]. A more accurate formulation of 25 would be 26:

26 /t/ followed by a high front vowel is realized as [č] after a continuant and as [š] elsewhere.

Returning to the form *right*, we see that the final consonant would be correctly determined as [č] rather than [š] if in the underlying representation there were a continuant preceding it – that is, if the underlying representation were /riφt/, where φ is some continuant. The continuant φ must, furthermore, be distinct from any of the continuants that actually appear phonetically in this position, namely the dental, labial, or palatal continuants in the unitalicized portion of *wrist*, *rift*, or *wished*. We may assume, then, that φ is the velar continuant /x/, which does not, of course, appear phonetically in English. The underlying form, then, would be /rixt/.

Consider now the derivation of *right.* By rule 23b, the representation /rixt/ becomes [rīt]. By rule 23c, the representation / rīt / becomes [rayt], which is the phonetic realization of *right.*

Consider next the derivation of *righteous.* Assuming that it has the same affix as *expeditious* and *repetitious*, we can represent it lexically as /rixtious/ (I do not concern myself here with the proper representation for *-ous*). Let us

suppose that the ordering of the rules so far discussed is the following: 23a, 24, 26, 23b, 23c, an ordering consistent with other relevant facts of English, given certain simplifications for convenience of exposition. Rule 23a is inapplicable and rule 24 is vacuous, when applied to the underlying form /rixtious/. Turning to rule 26, we see that it gives the form [rixčous]. Rule 23b now applies, giving [rīčous], and rule 23c gives [rayčous], which becomes [rayčəs] by reduction of unstressed vowels. Thus by rules 26 and 23, which are independently motivated, the underlying representation /rixt/ will be realized phonetically as [rayt] in isolation and as [rayč] in *righteous*, exactly as required.

These facts strongly suggest that the underlying phonological representation must be /rixt/ (in accord with the orthography and, of course, the history). A sequence of rules that must be in the grammar for other reasons gives the alternation *right-righteous.* Therefore, this alternation is not at all exceptional, but rather perfectly regular. Of course, the underlying representation is quite abstract; it is connected with the superficial phonetic shape of the signal only by a sequence of interpretive rules.

Putting the matter differently, suppose that a person knows English but does not happen to have the vocabulary item *righteous.* Hearing this form for the first time, he must assimilate it to the system he has learned. If he were presented with the derived form [rišəs], he would, of course, take the underlying representation to be exactly like that of *expedite*, *contrite*, and so on. But hearing [rayčəs], he knows that this representation is impossible; although the consonantal distinction [š]-[č] might easily be missed under ordinary conditions of language use, the vocalic distinction [i]-[ay] would surely be obvious. Knowing the rules of English and hearing the vocalic element [ay] instead of [i], he knows that either the form is a unique exception or it contains a sequence /i/ followed by velar and is subject to rule 26. The velar must be a continuant,[15] that is, /x/. But given that the velar is a continuant, it follows, if the form is regular (the null hypothesis, always), that the consonant must be [č], not [š], by rule 26. Thus, the hearer should perceive [rayčəs] rather than [rayšəs], even if the information as to the medial consonant is lacking in the received signal. Furthermore, the pressure to preserve regularity of alternations should act to block the superficial analogy to *expedite-expeditious* and *ignite-ignition*, and to preserve [č] as the phonetic realization of underlying /t/, as long as [ay] appears in place of expected [i], exactly as we observe to have occurred.

I do not mean this as a literal step-by-step account of how the form is learned, of course, but rather as a possible explanation of why the form resists a superficial (and in fact incorrect) analogy and preserves its status. We can explain the

[15] If it were a noncontinuant, it would have to be unvoiced, that is, /k/, since there are no voiced–voiceless consonant clusters in final position, by general rule. But it cannot be /k/, since /k/ remains in this position (for example, "direct," "evict," and so on).

perception and preservation in the grammar of the [č]-[š] contrast in *righteous-expeditious* on the basis of the perceived distinction between [ay] and [i] and the knowledge of a certain system of rules. The explanation rests on the assumption that the underlying representations are quite abstract, and the evidence cited suggests that this assumption is, in fact, correct.

A single example can hardly carry much conviction. A careful investigation of sound structure, however, shows that there are a number of examples of this sort, and that, in general, highly abstract underlying structures are related to phonetic representations by a long sequence of rules, just as on the syntactic level abstract deep structures are in general related to surface structures by a long sequence of grammatical transformations. Assuming the existence of abstract mental representations and interpretive operations of this sort, we can find a surprising degree of organization underlying what appears superficially to be a chaotic arrangement of data, and in certain cases we can also explain why linguistic expressions are heard, used, and understood in certain ways. One cannot hope to determine either the underlying abstract forms or the processes that relate them to signals by introspection; there is, furthermore, no reason why one should find this consequence in any way surprising.

The explanation sketched above is at the level of particular rather than universal grammar, as this distinction was formulated earlier. That is, we have accounted for a certain phenomenon on the basis of the assumption that certain rules appear in the internalized grammar, noting that these rules are, for the most part, independently motivated. Of course, considerations of universal grammar enter into this explanation insofar as they affect the choice of grammar on the basis of data. This interpenetration is unavoidable, as noted earlier. There are cases, however, where explicit principles of universal grammar enter more directly and clearly into a pattern of explanation. Thus, investigation of sound systems reveals certain very general principles of organization, some quite remarkable, governing phonological rules (see references in note 14). For example, it has been observed that certain phonological rules operate in a cycle, in a manner determined by the surface structure. Recall that the surface structure can be represented as a labeled bracketing of the utterance, such as 2. In English, the very intricate phonological rules that determine stress contours and vowel reduction apply to phrases bounded by paired brackets, in the surface structure, applying first to a minimal phrase of this sort, then to the next larger phrase, and so on, until the maximal domain of phonological processes is reached (in simple cases, the sentence itself). Thus, in the case of 2 the rules apply to the individual words (which, in a full description, would be assigned to categories and therefore bracketed), then to the phrases *a wise man* and *is honest*, and finally to the whole sentence. A few simple rules will give quite varied results, as the surface structures that determine their cyclic application vary.

Some simple effects of the principle of cyclic application are illustrated by such forms as those of 27:

27 a. *rel*a*xation*, *em*e*ndation*, *el*a*sticity*, *conn*e*ctivity*
b. *ill*u*stration*, *dem*o*nstration*, *dev*a*station*, *an*e*cdotal*

The unitalicized vowels are reduced to [ə] in 27b, but they retain their original quality in 27a. In some cases, we can determine the original quality of the reduced vowels of 27b from other derived forms (for example, *illustrative*, *demonstrative*). The examples of 27a differ from those of 27b morphologically in that the former are derived from underlying forms (namely, *relax*, *emend*, *elastic*, *connective*) that contain primary stress on the unitalicized vowel when these underlying forms appear in isolation; those of 27b do not have this property. It is not difficult to show that vowel reduction in English, the replacement of a vowel by [ə], is contingent upon lack of stress. We can therefore account for the distinction between 27a and 27b by assuming the cyclic principle just formulated. In the case of 27a, on the first, innermost cycle, stress will be assigned by general rules to the unitalicized vowels. On the next cycle, stress is shifted,[16] but the abstract stress assigned on the first cycle is sufficient to protect the vowel from reduction. In the examples of 27b, earlier cycles never assign an abstract stress to the unitalicized vowel, which thus reduces. Observe that it is an *abstract* stress that protects the vowel from reduction. The actual, phonetic stress on the unitalicized nonreduced vowels is very weak; it would be stress 4, in the usual convention. In general, vowels with this weak a phonetic stress reduce, but in this case the abstract stress assigned in the earlier cycle prevents reduction. Thus, it is the abstract underlying representation that determines the phonetic form, a primary role being played by the abstract stress that is virtually eliminated in the phonetic form.

In this case, we can provide an explanation for a certain aspect of perception and articulation in terms of a very general abstract principle, namely the principle of cyclic application of rules (see page 38). It is difficult to imagine how the language learner might derive this principle by "induction" from the data presented to him. In fact, many of the effects of this principle relate to perception and have little or no analogue in the physical signal itself, under normal conditions of language use, so that the phenomena on which the induction would have been based cannot be part of the experience of one who is not already making use of the principle. In fact, there is no procedure of induction or association that offers any hope of leading from such data as is available to a principle of this sort (unless, begging the question, we introduce the principle of cyclic application into the "inductive procedure" in some manner). Therefore, the conclusion

[16] In "connectivity," it is on the third cycle that the stress is shifted. The second cycle merely reassigns stress to the same syllable that is stressed on the first cycle.

seems warranted that the principle of cyclic application of phonological rules is an innate organizing principle of universal grammar that is used in determining the character of linguistic experience and in constructing a grammar that constitutes the acquired knowledge of language. At the same time, this principle of universal grammar offers an explanation for such phenomena as were noted in 27.

There is some evidence that a similar principle of cyclic application applies also on the syntactic level. John Ross has presented an ingenious analysis of some aspects of English pronominalization illustrating this.[17] Let us assume that pronominalization involves a process of "deletion" analogous to those processes discussed earlier in connection with examples 8– 18. This process, to first approximation, replaces one of two identical noun phrases by an appropriate pronoun. Thus, the underlying structure 28 will be converted into 29, by pronominalization.

28 John learned that John had won.

29 John learned that he had won.

Abstracting away from properties of 28 that are not essential to this discussion, we can present it in the form 30, where *x* and *y* are the identical noun phrases and *y* is the one pronominalized, and where the brackets bound sentential expressions.

30 $[\ldots x \ldots [\ldots y \ldots]]$

Notice that we cannot form 31 from 28 by pronominalization:[18]

31 He learned that John had won.

That is, we cannot have pronominalization in the case that would be represented as 32, using the conventions of 30:

32 $[\ldots y \ldots [\ldots x \ldots]]$

Consider next the sentences of 33:

33 a. That John won the race surprised him.
$[[\ldots x \ldots] \ldots y \ldots]$

b. John's winning the race surprised him.
$[[\ldots x \ldots] \ldots y \ldots]$

[17] J. Ross, "On the Cyclic Nature of English Pronominalization," in *To Honor Roman Jakobson* (New York: Humanities, 1967).

[18] Of course, 31 is a sentence, but "he" in the sentence does not refer to John as it does in 29. Thus, 31 is not formed by pronominalization if the two occurrences of "John" are intended to be different in reference. We exclude this case from discussion here. For some remarks bearing on this problem, see my *Aspects*, pp. 144–47.

c. That he won the race surprised John.
$[[\ldots y \ldots] \ldots x \ldots]$

d. His winning the race surprised John.
$[[\ldots y \ldots] \ldots x \ldots]$

Continuing with the same conventions, the forms are represented underneath, in each case. Summarizing, we see that of the possible types 30, 32, 33a, b and 33c, d, all permit pronominalization except 32. These remarks belong to the particular grammar of English.

Notice that alongside 33d we also have sentence 34:

34 Winning the race surprised John.

Given the framework we have been assuming throughout, 34 must be derived from the structure "John's winning the race surprised John." Hence, in this case pronominalization can be a full deletion.

Consider now sentences 35 and 36:

35 Our learning that John had won the race surprised him.

36 Learning that John had won the race surprised him.

Sentence 35 can be understood with "him" referring to John, but 36 cannot. Thus, 35 can be derived by pronominalization from 37, but 36 is not derived from 28:

37 [[Our learning [that John had won the race]] surprised John.]

38 [[John's learning [that John had won the race]] surprised John.]

What might be the explanation for this phenomenon? As Ross observes, it can be explained in terms of the particular grammar of English if we assume, in addition, that certain transformations apply in a cycle, first to innermost phrases, then to larger phrases, and so on – that is, if we assume that these transformations apply to the deep structure by a process analogous to the process by which phonological rules apply to the surface structure.[19] Making this assumption, let

[19] That transformational rules may be supposed to function in this way, itself a nontrivial fact if true, is suggested in my *Aspects*, Chapter 3. Ross's observation suggests that this principle of application is not only possible but also necessary. Other interesting arguments to this effect are presented in R. Jacobs and P. S. Rosenbaum, eds., *Readings in English Transformational Grammar*, Chapter 28. The matter is far from settled. In general, understanding of syntactic structure is much more limited than that of phonological structure, descriptions are much more rudimentary, and, correspondingly, principles of universal syntax are much less firmly established than principles of universal phonology, though the latter, needless to say, must also be regarded as tentative. In part, this may be due to the inherent complexity of the subject matter.

us consider the underlying structure 38. On the innermost cycle, pronominalization does not apply at all, since there is no second noun phrase identical to "John" in the most deeply embedded proposition. On the second cycle, we consider the phrase "[John's learning [that John had won the race]]." This can be regarded as a structure of the form 30, giving 39 by pronominalization; it cannot be regarded as of the form 32, giving 40 by pronominalization, because the particular grammar of English does not permit pronominalization in the case of 32, as we have noted:

39 John's learning [that he had won the race]

40 his learning [that John had won the race]

But 40 would have to be the form underlying 36. Hence, 36 cannot be derived by pronominalization from 38, although 35 can be derived from 37.

In this case, then, a principle of universal grammar combines with an independently established rule of particular English grammar to yield a certain rather surprising empirical consequence, namely that 35 and 36 must differ in the referential interpretation of the pronoun "him." Once again, as in the formally somewhat analogous case of vowel reduction discussed earlier in connection with examples 27a and 27b, it is quite impossible to provide an explanation in terms of "habits" and "dispositions" and "analogy." Rather, it seems that certain abstract and in part universal principles governing human mental faculties must be postulated to explain the phenomena in question. If the principle of cyclic application is indeed a regulative principle determining the form of knowledge of language for humans, a person who has learned the particular rules governing pronominalization in English would know, intuitively and without instruction or additional evidence, that 35 and 36 differ in the respect just noted.

The most challenging theoretical problem in linguistics is that of discovering the principles of universal grammar that interweave with the rules of particular grammars to provide explanations for phenomena that appear arbitrary and chaotic. Probably the most persuasive examples at this time (and also the most important ones, in that the principles involved are highly abstract and their operations quite intricate) are in the domain of phonology, but these are too complex to present within the scope of this lecture.[20] Another syntactic example

In part, it results from the fact that universal phonetics, which provides a kind of "empirical control" for phonological theory, is much more firmly grounded than universal semantics, which should, in principle, provide a partially analogous control for syntactic theory. In modern linguistics, phonetics (and, in part, phonology) has been studied in considerable depth and with much success, but the same cannot be said as yet for semantics, despite much interesting work.

[20] See references in note 14. The issue is discussed in a general way in my "Explanatory Models in Linguistics," in E. Nagel, P. Suppes, and A. Tarski, eds., *Logic, Methodology, and Philosophy of Science* (Stanford, Calif.: Stanford University Press, 1962); in my *Current Issues*, Section 2; in my *Aspects*, Chapter 1; and in other publications referred to in these references.

that illustrates the general problem in a fairly simple way is provided by the rules for formation of *wh*-questions in English.[21]

Consider such sentences as the following:

41
a. Who expected Bill to meet Tom?
b. Who(m) did John expect to meet Tom?
c. Who(m) did John expect Bill to meet?
d. What (books) did you order John to ask Bill to persuade his friends to stop reading?

As examples a, b, and c show, a noun phrase in any of the three italicized positions in a sentence such as *"John* expected *Bill* to meet *Tom"* can be questioned. The process is essentially this:

42
a. *wh-placement:* assign the marker *wh-* to a noun phrase.
b. *wh-inversion:* place the marked noun phrase at the beginning of the sentence.
c. *auxiliary attraction:* move a part of the verbal auxiliary or the copula to the second position in the sentence.
d. *phonological interpretation:* replace the marked noun phrase by an appropriate interrogative form.[22]

All four of these processes apply nonvacuously in the case of 41b and 41c. Sentence 41b, for example, is formed by applying *wh-placement* to the noun phrase "someone" in "John expected someone to meet Tom." Application of the process of *wh-inversion* (42b) gives "*wh*-someone John expected to meet Tom." The process of *auxiliary attraction* (42c) gives "*wh*-someone did John expect to meet Tom." Finally, the process of *phonological interpretation* (42d) gives 41b. Sentence 41d illustrates the fact that these processes can extract a noun phrase that is deeply embedded in a sentence – without limit, in fact.

Of the processes listed in 42, all but auxiliary attraction apply as well in the formation of relative clauses, giving such phrases as "the man who(m) John expected to meet Tom," and so on.

[21] This matter is discussed in my *Current Issues.* There are several versions of this monograph. The first, presented at the International Congress of Linguistics, 1962, appears in the *Proceedings of the Congress* with the title of the session at which it was presented, "Logical Basis of Linguistic Theory," ed. H. Lunt (New York: Humanities, 1964); a second appears in J. Fodor and J. J. Katz, eds., *Structure of Language: Readings in the Philosophy of Language* (Englewood Cliffs, N.J.: Prentice-Hall, 1964); the third, as a separate monograph (New York: Humanities, 1965). These versions differ in the treatment of the examples discussed here; none of the treatments is satisfactory, and the general problem remains open. New and interesting ideas on this matter are presented in J. Ross, "Constraints on Variables in Syntax," MIT doctoral thesis (unpublished). I follow here the general lines of the earliest of the three versions of *Current Issues*, which, in retrospect, seems to me to offer the most promising approach of the three.

[22] Actually, it seems that only indefinite singular noun phrases can be questioned (that is, "someone," "something," and so on), a fact that relates to the matter of recoverability of deletion mentioned in note 11. See my *Current Issues* for some discussion.

Notice, however, that there are certain restrictions on the formation of questions and relatives in this manner. Consider, for example, the sentences of 43:

43
a. For him to understand *this lecture* is difficult.
b. It is difficult for him to understand *this lecture.*
c. He read the book that interested *the boy.*
d. He believed the claim that John tricked *the boy.*
e. He believed the claim that John made about *the boy.*
f. They intercepted John's message to *the boy.*

Suppose that we try applying the processes of interrogative and relative formation to the italicized noun phrases in 43. We should derive the following interrogatives and relatives from 43a–43f, respectively:

44
aI. *What is for him to understand difficult?
aR. *a lecture that for him to understand is difficult
bI. What is it difficult for him to understand?
bR. a lecture that it is difficult for him to understand
cI. *Who did he read the book that interested?
cR. *the boy who he read the book that interested
dI. *Who did he believe the claim that John tricked?
dR. *the boy who he believed the claim that John tricked
eI. *Who did he believe the claim that John made about?
eR. *the boy who he believed the claim that John made about
fI. *Who did they intercept John's message to?
fR. *the boy who they intercepted John's message to

Of these, only bI and bR are fully acceptable, and cases a, c, d, and e are quite impossible, although it would be quite clear what they meant, were they grammatically permissible. It is not at all obvious how the speaker of English knows this to be so. Thus, sentences 43a and 43b are synonymous, yet only 43b is subject to the processes in question. And although these processes do not apply to 43d and 43f, they can be applied, with much more acceptable results, to the very similar sentences 45a and 45b:

45
a. He believed that John tricked *the boy.* (Who did he believe that John tricked? – the boy who he believed that John tricked)
b. They intercepted a message to *the boy.* (Who did they intercept a message to? – the boy who they intercepted a message to)

In some unknown way, the speaker of English devises the principles of 42 on the basis of data available to him; still more mysterious, however, is the fact that he knows under what formal conditions these principles are applicable. It can hardly be seriously maintained that every normal speaker of English has had his behavior "shaped" in the indicated ways by appropriate reinforcement. The sentences of 43, 44, and 45 are as "unfamiliar" as the vast majority of those

that we encounter in daily life, yet we know intuitively, without instruction or awareness, how they are to be treated by the system of grammatical rules that we have mastered.

It seems, once again, that there is a general principle that accounts for many such facts. Notice that in 43a the italicized noun phrase is contained within another noun phrase, namely "for him to understand this lecture," which is the subject of the sentence. In 43b, however, a rule of *extraposition* has placed the phrase "for him to understand this lecture" outside of the subject noun phrase, and in the resulting structure this phrase is not a noun phrase at all, so that the italicized phrase in 43b is no longer contained within a noun phrase. Suppose we were to impose on grammatical transformations the condition that no noun phrase can be extracted from within another noun phrase – more generally, that if a transformation applies to a structure of the form

$$\left[\text{S} \ldots [\text{A} \ldots] \text{A} \ldots \right] \text{S}$$

for any category A, then it must be so interpreted as to apply to the *maximal* phrase of the type A.[23] Then the processes of 42 would be blocked, as required, in cases 43a, c, d , e, and f, but not in 43b. We will return shortly to 45.

There are other examples that support a principle of this sort, which we will refer to as the *A-over-A* principle. Consider the sentences of 46:

46 a. John kept the car in *the garage.*
b. Mary saw the man walking toward *the railroad station.*

Each of these is ambiguous. Thus, 46a can mean that the car in the garage was kept by John, or that the car was kept in the garage by John. In the first case, the italicized phrase is part of a noun phrase, "the car in the garage"; in the latter case it is not. Similarly, 46b can mean that the man walking toward the railroad station was seen by Mary, or that the man was seen walking toward the railroad station by Mary (or, irrelevantly to this discussion, that Mary, while walking toward the railroad station, saw the man). Again, in the first case, the italicized phrase is part of a noun phrase, "the man walking toward the railroad station"; in the latter case, it is not. But now consider the two interrogatives of 47:

47 a. What (garage) did John keep the car in?
b. What did Mary see the man walking toward?

[23] We might extend this principle to the effect that this transformation must also apply to the *minimal* phrase of the type S (sentence). Thus, the sentence

$$\left[_{\text{S}} \text{John was convinced that } [_{\text{S}} \text{Bill would leave before dark}]_{\text{S}} \right]_{\text{S}}$$

can be transformed to "John was convinced that before dark Bill would leave" but not to "before dark John was convinced that Bill would leave," which must have a different source. Like the original principle, this extension is not without its problems, but it has a certain amount of support nevertheless.

Each of these is unambiguous and can have only the interpretation of the underlying sentence in which the italicized phrase is not part of another noun phrase. The same is true of the relatives formed from 46, and these facts too would be explained by the A-over-A principle. There are many similar examples.

A slightly more subtle case that might, perhaps, be explained along the same lines is provided by such sentences as 48 and 49:

48 John has the best proof of that theorem.

49 What theorem does John have the best proof of?

In its most natural interpretation, sentence 48 describes a situation in which a number of people have proofs of that theorem, and John's is the best. The sense thus suggests that "best" modifies the nominal phrase "proof of that theorem," which contains another nominal phrase, "that theorem."[24] The A-over-A principle would therefore imply that the phrase "that theorem" not be subject to the processes of 42. Hence, 49 would not be derived by these processes from 48. And, in fact, sentence 49 has an interpretation rather different from that of 48. Sentence 49 is appropriate to a situation in which John has proofs of a number of theorems, and the questioner is asking which of these proofs is the best. The underlying structure, whatever it may be, would associate "best" with "proof," not with "proof of that theorem," so that "that theorem" is not embedded within a phrase of the same type and is therefore subject to questioning (and, similarly, to relativization).

The general principle just proposed has a certain explanatory force, as such examples illustrate. If postulated as a principle of universal grammar, it can explain why the particular rules of English operate to generate certain sentences while rejecting others, and to assign sound–meaning relations in ways that appear, superficially, to violate regular analogies. Putting the matter in different terms, if we assume that the A-over-A principle is a part of the innate schematism that determines the form of knowledge of language, we can account for certain aspects of the knowledge of English possessed by speakers who obviously have not been trained and who have not even been presented with data bearing on the phenomena in question in any relevant way, so far as can be ascertained.

Further analysis of data of English reveals, not unexpectedly, that this account is oversimplified and runs up against many difficulties. Consider, for example, sentences 50 and 51:

[24] Space does not permit a discussion of the distinction implied here in the loose terminology, "noun phrase"–"nominal phrase," but this is not crucial to the point at issue. See my "Remarks on Nominalization," in R. Jacobs and P. S. Rosenbaum, eds., *Readings in English Transformational Grammar.* There are other interpretations of 49 (for example, with contrastive stress on "John"), and there are many open problems relating to such structures as these.

50 John thought (that) Bill had read *the book.*

51 John wondered why Bill had read *the book.*

In the case of 50, the italicized phrase is subject to interrogation and relativization, but not in the case of 51. It is unclear whether the phrases "that Bill had read the book" and "why Bill had read the book" are noun phrases. Suppose that they are not. Then sentence 50 is handled in accordance with the A-over-A principle, but not 51. To explain the blocking of the processes of 42 in the case of 51, we would have to assign the phrase "why Bill had read the book" to the same category as "the book." In fact, there is a natural suggestion along these lines. Sentence 51 is typical in that the phrase from which the noun phrase is to be extracted is itself a *wh-* phrase, rather than a *that-* phrase. Suppose that the process of *wh-* placement (42a) assigns the element *wh-* not only to "the book" in 51 but also to the proposition containing it. Thus, both "*wh-* the book" and "why Bill had read the book" belong to the category *wh-*, which would now be regarded as a syntactic feature of a sort discussed in my *Aspects of the Theory of Syntax*, Chapter 2 (see note 6). Under these assumptions, the A-over-A principle will serve to explain the difference between 50 and 51.

Suppose that the phrases in question are noun phrases. Now it is 50, not 51, that poses the problem. Assuming that our analysis is correct so far, there must be some rule that assigns to the proposition "that Bill had read the book" a property of "transparency" that permits noun phrases to be extracted from it even though it is a noun phrase. There are, in fact, other examples that suggest the necessity for such a rule, presumably a rule of the particular grammar of English. Thus, consider sentences 52, 53, and 54:

52 Who would you approve of my seeing?

53 What would you approve of John's drinking?

54 *What would you approve of John's excessive drinking of?

Sentences 52 and 53 are formed by applying the processes of interrogation to a noun phrase contained in the larger phrases "my seeing –," "John's drinking –." Hence, these larger noun phrases are transparent to the extraction operation. However, as 54 indicates, the italicized noun phrase in 55 is not transparent to this operation:

55 You would approve of John's excessive drinking of the beer.

These examples are typical of many that suggest what the rule might be that assigns transparency. Earlier we discussed sentence 56 (sentence 4), pointing out that it is ambiguous:

56 I disapprove of John's drinking.

Under one interpretation, the phrase "John's drinking" has the internal structure of a noun phrase. Thus, the rule that inserts adjectives (**3d**) between a determiner and a noun applies, giving "John's excessive drinking"; and, in fact, other determiners may replace "John's" – "the," "that," "much of that," and so on. Under this interpretation, the phrase "John's drinking" behaves exactly like "John's refusal to leave," "John's rejection of the offer," and so on. Under the other interpretation, "John's drinking (the beer)" does not have the internal structure of a noun phrase and is handled analogously to "John's having read the book," "John's refusing to leave," "John's rejecting the offer," and so on, none of which permits adjective insertion or replacement of "John's" by other determiners. Suppose that we postulate a rule of English grammar that assigns transparency, in the sense just defined, to noun phrases that are also propositions lacking the internal structure of noun phrases. Thus, the phrases "that Bill had read the book" in 50, "my seeing –" in the structure underlying 52, and "John's drinking –" in the structure underlying 53 would be assigned transparency; more precisely, the dominant noun phrase in these examples would not serve to block extraction by the A-over-A principle. In sentence 51, extraction would still be blocked by the category *wh*-, along the lines indicated earlier. And sentence 54 is ruled out because the relevant noun phrase of the underlying structure, "John's excessive drinking of –," does have the internal structure of a noun phrase, as just noted, and therefore is not subject to the special rule of English grammar that assigns transparency to the category NP when this category dominates a proposition that lacks the internal structure of an NP.

There are a few other cases that suggest the need for rules of particular grammar assigning transparency in this sense. Consider sentences 57 and 58:

57 a. They intercepted John's message to *the boy.* (Sentence 43f)
b. He saw John's picture of *Bill.*
c. He saw the picture of *Bill.*

58 a. They intercepted a message to *the boy.* (Sentence 45b)
b. He saw a picture of *Bill.*
c. He has a belief in *justice.*
d. He has faith in *Bill's integrity.*

The italicized noun phrases in 57 are not subject to the processes of interrogation and relativization, in accordance with the A-over-A principle, as we have already noted. In the case of 58, interrogation and relativization seem much more natural in these positions, at least in informal spoken English. Thus, the noun phrases containing the italicized phrases must be assigned transparency. It seems that what is involved is indefiniteness of the dominating noun phrase; if so, then for certain dialects there is a rule assigning transparency to a noun phrase of the form

59 [NP indefinite . . . NP]NP

There remain a number of very serious problems that seem to resist solution by such extensions and modifications of the A-over-A principle. Notice that this principle is formulated in a way that is not really well supported by the examples so far given. If the A-over-A principle were true in general, we would expect to find cases in which a phrase of category A cannot be extracted from a larger phrase of category A, for various choices of A. In fact, the examples given so far involve only A = noun phrase (or, perhaps, A = [+ *wh*-], as in the discussion of 51). Hence, an alternative formulation of the principle consistent with the facts just noted would assign nontransparency as an ad hoc property of certain types of noun phrases (and perhaps other constructions), rather than as a property of a category A dominating another category of the type A. Given just the facts so far presented, it would be proper to postulate the A-over-A principle instead of this alternative precisely because the A-over-A principle has a certain naturalness, whereas the alternative is entirely ad hoc, a listing of nontransparent structures. But there is crucial evidence, pointed out by John Ross (see reference in note 21), suggesting that the A-over-A principle is not correct. Ross points out that in the constructions from which noun phrases cannot be extracted, adjectives also cannot be extracted. Thus, consider the contexts "I believe that John saw –," "I believe the claim that John saw –," and "I wonder whether John saw –." From the first of these, but not the second or third, we can extract a noun phrase in interrogation or relativization, a fact that we have been attempting to account for by modifications of the A-over-A principle. But the same is true of extraction of adjectives. Thus we can form "handsome though I believe that John is," but not *"handsome though I believe the claim that John is," *"handsome though I wonder whether John is," etc. Whether one can extend the approach just discussed to account for this problem in some natural way, I do not know; at the moment, I see no approach that does not involve a perfectly ad hoc step. Perhaps this indicates that the approach through the A-over-A principle is incorrect, leaving us for the moment with only a collection of constructions in which extraction is, for some reason, impossible to accomplish.

Whatever the answer will prove to be, the complex of problems just discussed is a typical and important illustration of the kind of topic that is at the border of research today, in the sense mentioned at the outset of this lecture: that is, certain problems can be formulated clearly within a framework of ideas that is reasonably clear and well understood; certain partial solutions can be advanced; and a range of examples can be discovered where these solutions fail, leaving open for the time being the question whether what is needed is further elaboration and sharpening or a radically different approach.

I have so far discussed several kinds of conditions that transformations must meet: conditions of deletion, of the sort brought out by examples 8–18; the principle of cyclic application, illustrated by the discussion of examples 28–40 (with the phonological analogue discussed in connection with 27); and the

A-over-A principle that was proposed as the basis for an explanation of such phenomena as are illustrated by examples 44–58. In each case, there is some reason to believe that the principle is appropriate, though there is no lack of evidence showing that the principle is inadequately formulated or, perhaps, misconceived. As a final illustration of this state of affairs, typical of the borderline of research that exists in linguistics as in any other field, consider a problem first discussed by Peter Rosenbaum (see reference in note 6). Consider the sentences of 60:

60 a. John agreed to go.
b. John persuaded Bill to leave.
c. Finding Tom there caused Bill to wonder about John.

In interpreting these sentences, we supply a "missing subject" for the verbs "go," "leave," and "find," respectively. In 60a, we understand the subject of "go" to be "John"; in 60b, we understand the subject of "leave" to be "Bill"; in 60c, we understand the subject of "find" and the subject of "wonder" to be "Bill." In terms of the framework presupposed so far, it would be natural (though perhaps not necessary, as we will see below) to regard this missing subject as the actual subject in the deep structure, eliminated by a deletion operation. Thus, the underlying deep structures might be something like 61:

61 a. John agreed [John go]
b. John persuaded Bill [Bill leave]
c. [Bill find Tom there] caused Bill to wonder about John

On the other hand, the facts indicate clearly that the sentences of 60 cannot derive from, say, 62:

62 a. John agreed [someone go]
b. John persuaded Bill [John leave]
c. [John find Tom there] caused Bill to wonder about John

It would be difficult to argue that in such cases there is an intrinsic semantic consideration ruling against such structures as 62. For example, one might interpret 62a as meaning that John agreed that someone should go; 62b as meaning that John persuaded Bill that he (John) would (should) leave; 62c as meaning that John's finding Tom there caused Bill to wonder about John. There must be some general syntactic principle that rules against 62 as possible sources for 60 and that causes us to interpret 60 as based rather on 61. Rosenbaum suggests that what is involved is a certain condition on deletion operations, an "erasure principle" that prescribes roughly that the subject of an embedded proposition is deleted by the nearest noun phrase outside of this proposition, "nearness" being measured in terms of the number of branches in a representation such as

1′ or 2′.[25] As he shows, a great many examples of varied sorts can be explained on this general assumption, which, like the others that I have been reviewing, involves a condition on transformations that would constitute part of universal grammar.

Here too, however, certain problems arise. Consider, for example, the following cases:[26]

63 John promised Bill to leave.

64 a. John gave me the impression of working on that problem.
b. John gave me the suggestion of working on that problem.

65 a. John asked me what to wear.
b. John told me what to wear.

66 John asked Bill for permission to leave.

67 a. John begged Bill to permit him to stay.
b. John begged Bill to be permitted to stay.
c. John begged Bill to be shown the new book.

68 John made an offer to Bill (received advice from Bill, received an invitation from Bill) to stay.

69 John helped Bill write the book.

Sentence 63 violates the principle, since it is John, not Bill, who is to leave. In 64a, "John" is understood to be the subject of "work," whereas in the apparently analogous sentence 64b the subject is understood to be "I." In the case of 65a, it is "John" that is the understood subject of "wear"; in 65b, it is "I." In the case of 66, "John" is the understood subject of "leave" and "Bill" of "permit," underlying "permission," presumably; in the case of 67a, "Bill" is the understood surface subject of the embedded proposition, but in 67b and 67c it is "John," although "Bill" is the "nearest" noun phrase in all three cases, in Rosenbaum's sense. In 68, it is "John" that is understood as the subject of "stay," in apparent contradiction to the principle, though much depends on unresolved questions as to how these sentences are to be analyzed. The case of 69 is obscure in other ways. The erasure principle would suggest that "Bill" is the subject of "write," although of course the sentence does not imply that Bill wrote the book – rather John and Bill did, together. But there is a difficulty in pursuing this interpretation. Thus, from 69 we can conclude that John helped

[25] In yet-unpublished work, David Perlmutter has presented a strong argument that what is involved is not a condition on transformations but rather a condition on well-formed deep structures. The distinction is not crucial for what follows but would become important at a less superficial level of discussion.

[26] Examples 63 and 67 are discussed by Rosenbaum; 64 was pointed out by Maurice Gross; 65 was pointed out in a different connection by Zeno Vendler, "Nominalizations," in *Transformations and Discourse Analysis Papers*, No. 55 (Philadelphia: University of Pennsylvania, 1964), p. 67.

write the book, but from the apparently analogous sentence "John helped the cat have kittens," we cannot deduce that "John helped have kittens," which is deviant, a fact that suggests that somehow there must be a grammatical relation between "John" and "write" in 69. To put it differently, the problem is how to account for "John helped write the book" as analogous to 60a, since obviously the analogue to 61a won't do as a source.

Without pursuing the matter any further, we can see that although the erasure principle has much to recommend it and is probably somehow involved in the correct solution to this network of problems, there is much evidence still to be accounted for. As in the other cases mentioned, there are a variety of problems relating to the conditions that determine applicability of transformations, problems that still resist any near-definitive solution, though some interesting and illuminating proposals can be made that seem to go part of the way toward a general solution.

In discussing the nature of grammatical operations, I have restricted myself to syntactic and phonological examples, avoiding questions of semantic interpretation. If a grammar is to characterize the full linguistic competence of the speaker–hearer, it must comprise rules of semantic interpretation as well, but little is known of any depth regarding this aspect of grammar. In the references cited earlier (see note 6), it is proposed that a grammar consists of a syntactic component that specifies an infinite set of paired deep and surface structures and expresses the transformational relationship between these paired elements, a phonological component that assigns a phonetic representation to the surface structure, and a semantic component that assigns a semantic representation to the deep structure. As noted earlier (p. 27; see also pp. 94–97), I think there is strong evidence that aspects of the surface structure are also relevant to semantic interpretation.[27] However this may be, there can be little doubt that a full grammar must contain fairly intricate rules of semantic interpretation, keyed, at least in part, to fairly specific properties of the lexical items and formal structures of the language in question. To mention just one example, consider sentence 70:

70 John has lived in Princeton.

From the assumption that this sentence has been properly used to make a statement, we can conclude that John is a person (one would not say that his dog has lived in Princeton); that Princeton is a place meeting certain physical and

[27] For some remarks concerning this problem, see my "Surface Structure and Semantic Interpretation," in R. Jakobson, ed., *Studies in General and Oriental Linguistics* (Tokyo: TEC Corporation for Language and Educational Research, 1970). Literature on semantic interpretation of syntactic structures is expanding fairly rapidly. For recent discussion, see J. J. Katz, *The Philosophy of Language* (New York: Harper & Row, 1966); U. Weinreich, "Explorations in Semantic Theory," in T. A. Sebeok, ed., *Current Trends in Linguistics*, Vol. III (New York: Humanities, 1966); J. J. Katz, "Recent Issues in Semantic Theory," in *Foundations of Language*, Vol. 3, No. 2, May 1967, pp. 124–94; and many other papers.

sociological conditions (given that "Princeton" is a proper noun); that John is now alive (I can say that I have lived in Princeton, but I cannot now say "Einstein has lived in Princeton" – rather, "Einstein lived in Princeton"); and so on. The semantic interpretation of 70 must be such as to account for these facts.

In part, such questions as these might be subsumed under a still-to-be-developed universal semantics, in which concepts and their relations are analyzed in a very general way; to take a classical example, it might be argued that the relation of meaning between "John is proud of what Bill did" and "John has some responsibility for Bill's actions" should be explained in terms of the universal concepts of pride and responsibility, just as on the level of sound structure one might appeal to a principle of universal phonetics to account for the fact that when a velar consonant becomes palatal it ordinarily becomes strident (see references in note 14, for discussion). The proposal looks less attractive when applied to the case of 70, for example, with respect to the fact that proper use of 70 implies that John is now alive. When we try to pursue such questions, we soon become lost in a tangle of confused issues and murky problems, and it is difficult to propose answers that carry any conviction. For this reason, I am unable to discuss conditions on rules of semantic interpretation that might be analogous to the conditions on syntactic and phonological rules mentioned earlier.

Observe that I might well have been mistaken in the preceding remarks in assuming that the topics discussed belong to syntax rather than to the semantic component of a grammar, or to some domain in which semantic and syntactic rules interpenetrate. The issues are too clouded for us to be able to say that this is an empirical question, as matters now stand; but when they are sharpened, we may find that an empirical question can be posed. Consider, for example, the discussion of the erasure principle in syntax. Joseph Emonds has suggested (in unpublished work) that it is incorrect to assume, as I did, that the sentences of 60 are interpreted through reference to the underlying structures of 61. Rather, he argues that what I took to be the embedded proposition has no subject at all in the underlying form generated by the syntactic component, and a general rule of semantic interpretation takes the place of Rosenbaum's erasure principle. Whether this is correct I do not know, but it is certainly a possibility. We can expect, as research continues into problems of grammar, that the boundaries that seem clear today may shift in unpredictable ways, or that some new basis for the organization of grammar may replace the framework that now seems appropriate.

The conditions on grammatical rules that I have been discussing are complex and only partially understood. It should be emphasized, however, that even some of the simplest and clearest conditions of the form of grammar are in no sense necessary properties of a system that fulfills the functions of human language. Correspondingly, the fact that they hold true of human languages in general and

play a role in the acquired linguistic competence of the speaker–hearer cannot be lightly dismissed. Consider, for example, the simple fact that grammatical transformations are invariably *structure-dependent* in the sense that they apply to a string of words[28] by virtue of the organization of these words into phrases. It is easy to imagine *structure-independent* operations that apply to a string of elements quite independently of its abstract structure as a system of phrases. For example, the rule that forms the interrogatives of 71 from the corresponding declaratives of 72 (see note 10) is a structure-dependent rule interchanging a noun phrase with the first element of the auxiliary.

71 a. Will the members of the audience who enjoyed the play stand?
b. Has Mary lived in Princeton?
c. Will the subjects who will act as controls be paid?

72 a. The members of the audience who enjoyed the play will stand.
b. Mary has lived in Princeton.
c. The subjects who will act as controls will be paid.

In contrast, consider the operation that inverts the first and last words of a sentence, or that arranges the words of a sentence in increasing length in terms of phonetic segments ("alphabetizing" in some specified way for items of the same length), or that moves the left-most occurrence of the word "will" to the extreme left – call these O_1, O_2, and O_3, respectively. Applying O_1 to 72a, we derive 73a; applying O_2 to 72b, we derive 73b; applying O_3 to 72c, we derive 73c:

73 a. stand the members of the audience who enjoyed the play will
b. in has lived Mary Princeton
c. will the subjects who act as controls will be paid

The operations O_1, O_2, and O_3 are structure-independent. Innumerable other operations of this sort can be specified.

There is no a priori reason why human language should make use exclusively of structure-dependent operations, such as English interrogation, instead of structure-independent operations, such as O_1, O_2, and O_3. One can hardly argue that the latter are more "complex" in some absolute sense; nor can they be shown to be more productive of ambiguity or more harmful to communicative efficiency. Yet no human language contains structure-independent operations among (or replacing) the structure-dependent grammatical transformations. The language-learner knows that the operation that gives 71 is a possible candidate for a grammar, whereas O_1, O_2, and O_3, and any operations like them, need not be considered as tentative hypotheses.

[28] More properly, to a string of minimal linguistic units that may or may not be words.

If we establish the proper "psychic distance" from such elementary and commonplace phenomena as these, we will see that they really pose some nontrivial problems for human psychology. We can speculate about the reason for the reliance on structure-dependent operations,[29] but we must recognize that any such speculation must involve assumptions regarding human cognitive capacities that are by no means obvious or necessary. And it is difficult to avoid the conclusion that whatever its function may be, the reliance on structure-dependent operations must be predetermined for the language-learner by a restrictive initial schematism of some sort that directs his attempts to acquire linguistic competence. Similar conclusions seem to me warranted, a fortiori, in the case of the deeper and more intricate principles discussed earlier, whatever their exact form may turn out to be.

To summarize: along the lines that have been outlined here, we might develop on the one hand a system of general principles of universal grammar,[30] and on the other, particular grammars that are formed and interpreted in accordance with these principles. The interplay of universal principles and particular rules leads to empirical consequences such as those we have illustrated; at various levels of depth, these rules and principles provide explanations for facts about linguistic competence – the knowledge of language possessed by each normal speaker – and about some of the ways in which this knowledge is put to use in the performance of the speaker or hearer.

The principles of universal grammar provide a highly restrictive schema to which any human language must conform, as well as specific conditions determining how the grammar of any such language can be used. It is easy to imagine alternatives to the conditions that have been formulated (or those that are often tacitly assumed). These conditions have in the past generally escaped notice, and we know very little about them today. If we manage to establish the appropriate "psychic distance" from the relevant phenomena and succeed in "making them strange" to ourselves, we see at once that they pose very serious problems that cannot be talked or defined out of existence. Careful consideration of such problems as those sketched here indicates that to account for the normal

[29] See G. A. Miller and N. Chomsky, "Finitary Models of Language Users, Part II," in R. D. Luce, R. Bush, and E. Galanter, eds., *Handbook of Mathematical Psychology*, Vol. II (New York: Wiley, 1963), for some proposals regarding this matter.

[30] Notice that we are interpreting "universal grammar" as a system of conditions on grammars. It may involve a skeletal substructure of rules that any human language must contain, but it also incorporates conditions that must be met by such grammars and principles that determine how they are interpreted. This formulation is something of a departure from a traditional view that took universal grammar to be simply a substructure of each particular grammar, a system of rules at the very core of each grammar. This traditional view has also received expression in recent work. It seems to me to have little merit. As far as information is available, there are heavy constraints on the form and interpretation of grammar at all levels, from the deep structures of syntax, through the transformational component, to the rules that interpret syntactic structures semantically and phonetically.

use of language we must attribute to the speaker–hearer an intricate system of rules that involve mental operations of a very abstract nature, applying to representations that are quite remote from the physical signal. We observe, furthermore, that knowledge of language is acquired on the basis of degenerate and restricted data and that it is to a large extent independent of intelligence and of wide variations in individual experience.

If a scientist were faced with the problem of determining the nature of a device of unknown properties that operates on data of the sort available to a child and gives as "output" (that is, as a "final state of the device," in this case) a particular grammar of the sort that it seems necessary to attribute to the person who knows the language, he would naturally search for inherent principles of organization that determine the form of the output on the basis of the limited data available. There is no reason to adopt a more prejudiced or dogmatic view when the device of unknown properties is the human mind; specifically, there is no reason to suppose, in advance of any argument, that the general empiricist assumptions that have dominated speculation about these matters have any particular privileged claim. No one has succeeded in showing why the highly specific empiricist assumptions about how knowledge is acquired should be taken seriously. They appear to offer no way to describe or account for the most characteristic and normal constructions of human intelligence, such as linguistic competence. On the other hand, certain highly specific assumptions about particular and universal grammar give some hope of accounting for the phenomena that we face when we consider knowledge and use of language. Speculating about the future, it seems not unlikely that continued research along the lines indicated here will bring to light a highly restrictive schematism that determines both the content of experience and the nature of the knowledge that arises from it, thus vindicating and elaborating some traditional thinking about problems of language and mind. It is to this matter, among others, that I shall turn in the final lecture.

3 Linguistic contributions to the study of mind: future

In discussing the past, I referred to two major traditions that have enriched the study of language in their separate and very different ways; and in my last lecture, I tried to give some indication of the topics that seem on the immediate horizon today, as a kind of synthesis of philosophical grammar and structural linguistics begins to take shape. Each of the major traditions of study and speculation that I have been using as a point of reference was associated with a certain characteristic approach to the problems of mind; we might say, without distortion, that each evolved as a specific branch of the psychology of its time, to which it made a distinctive contribution.

It may seem a bit paradoxical to speak of structural linguistics in this way, given its militant anti-psychologism. But the paradox is lessened when we take note of the fact that this militant anti-psychologism is no less true of much of contemporary psychology itself, particularly of those branches that until a few years ago monopolized the study of use and acquisition of language. We live, after all, in the age of "behavioral science," not of "the science of mind." I do not want to read too much into a terminological innovation, but I think that there is some significance in the ease and willingness with which modern thinking about man and society accepts the designation "behavioral science." No sane person has ever doubted that behavior provides much of the evidence for this study – all of the evidence, if we interpret "behavior" in a sufficiently loose sense. But the term "behavioral science" suggests a not-so-subtle shift of emphasis toward the evidence itself and away from the deeper underlying principles and abstract mental structures that might be illuminated by the evidence of behavior. It is as if natural science were to be designated "the science of meter readings." What, in fact, would we expect of natural science in a culture that was satisfied to accept this designation for its activities?

Behavioral science has been much preoccupied with data and organization of data, and it has even seen itself as a kind of technology of control of behavior. Anti-mentalism in linguistics and in philosophy of language conforms to this shift of orientation. As I mentioned in my first lecture, I think that one major indirect contribution of modern structural linguistics results from its success in making explicit the assumptions of an anti-mentalistic, thoroughly operational

and behaviorist approach to the phenomena of language. By extending this approach to its natural limits, it laid the groundwork for a fairly conclusive demonstration of the inadequacy of any such approach to the problems of mind.

More generally, I think that the long-range significance of the study of language lies in the fact that in this study it is possible to give a relatively sharp and clear formulation of some of the central questions of psychology and to bring a mass of evidence to bear on them. What is more, the study of language is, for the moment, unique in the combination it affords of richness of data and susceptibility to sharp formulation of basic issues.

It would, of course, be silly to try to predict the future of research, and it will be understood that I do not intend the subtitle of this lecture to be taken very seriously. Nevertheless, it is fair to suppose that the major contribution of the study of language will lie in the understanding it can provide as to the character of mental processes and the structures they form and manipulate. Therefore, instead of speculating on the likely course of research into the problems that are coming into focus today,[1] I will concentrate here on some of the issues that arise when we try to develop the study of linguistic structure as a chapter of human psychology.

It is quite natural to expect that a concern for language will remain central to the study of human nature, as it has been in the past. Anyone concerned with the study of human nature and human capacities must somehow come to grips with the fact that all normal humans acquire language, whereas acquisition of even its barest rudiments is quite beyond the capacities of an otherwise intelligent ape – a fact that was emphasized, quite correctly, in Cartesian philosophy.[2] It is widely thought that the extensive modern studies of animal communication challenge this classical view; and it is almost universally taken

[1] A number of such problems might be enumerated – for example, the problem of how the intrinsic content of phonetic features determines the functioning of phonological rules, the role of universal formal conditions in restricting the choice of grammars and the empirical interpretation of such grammars, the relations of syntactic and semantic structure, the nature of universal semantics, performance models that incorporate generative grammars, and so on.

[2] Modern attempts to train apes in behavior that the investigators regard as language-like confirm this incapacity, though it may be that the failures are to be attributed to the technique of operant conditioning and therefore show little about the animal's actual abilities. See, for example, the report by C. B. Ferster, "Arithmetic Behavior in Chimpanzees," in *Scientific American*, May 1964, pp. 98–106. Ferster attempted to teach chimpanzees to match the binary numbers 001, . . . , 111 to sets of one to seven objects. He reports that hundreds of thousands of trials were required for 95 percent accuracy to be achieved, even in this trivial task. Of course, even at this stage the apes had not learned the principle of binary arithmetic; they would not, for example, be able to match a four-digit binary number correctly, and, presumably, they would have done just as badly in the experiment had it involved an arbitrary association of the binary numbers to sets rather than the association determined by the principle of the binary notation. Ferster overlooks this crucial point and therefore concludes, mistakenly, that he has taught the rudiments of symbolic behavior. The confusion is compounded by his definition of language as "a set of symbolic stimuli that control behavior" and by his strange belief that the "effectiveness" of language arises from the fact that utterances "control almost identical performances in speaker and listener."

for granted that there exists a problem of explaining the "evolution" of human language from systems of animal communication. However, a careful look at recent studies of animal communication seems to me to provide little support for these assumptions. Rather, these studies simply bring out even more clearly the extent to which human language appears to be a unique phenomenon, without significant analogue in the animal world. If this is so, it is quite senseless to raise the problem of explaining the evolution of human language from more primitive systems of communication that appear at lower levels of intellectual capacity. The issue is important, and I would like to dwell on it for a moment.

The assumption that human language evolved from more primitive systems is developed in an interesting way by Karl Popper in his recently published Arthur Compton Lecture, "Clouds and Clocks." He tries to show how problems of freedom of will and Cartesian dualism can be solved by the analysis of this "evolution." I am not concerned now with the philosophical conclusions that he draws from this analysis, but with the basic assumption that there is an evolutionary development of language from simpler systems of the sort that one discovers in other organisms. Popper argues that the evolution of language passed through several stages, in particular a "lower stage" in which vocal gestures are used for expression of emotional state, for example, and a "higher stage" in which articulated sound is used for expression of thought – in Popper's terms, for description and critical argument. His discussion of stages of evolution of language suggests a kind of continuity, but in fact he establishes no relation between the lower and higher stages and does not suggest a mechanism whereby transition can take place from one stage to the next. In short, he gives no argument to show that the stages belong to a single evolutionary process. In fact, it is difficult to see what links these stages at all (except for the metaphorical use of the term "language"). There is no reason to suppose that the "gaps" are bridgeable. There is no more of a basis for assuming an evolutionary development of "higher" from "lower" stages, in this case, than there is for assuming an evolutionary development from breathing to walking; the stages have no significant analogy, it appears, and seem to involve entirely different processes and principles.

A more explicit discussion of the relation between human language and animal communication systems appears in a recent discussion by the comparative ethologist W. H. Thorpe.[3] He points out that mammals other than man appear to lack the human ability to imitate sounds, and that one might therefore have expected birds (many of which have this ability to a remarkable extent) to

[3] W. H. Thorpe, "Animal Vocalization and Communication," in F. L. Darley, ed., *Brain Mechanisms Underlying Speech and Language* (New York: Grune and Stratton, 1967), pp. 2–10 and the discussions on pp. 19 and 84–85.

be "the group which ought to have been able to evolve language in the true sense, and not the mammals." Thorpe does not suggest that human language "evolved" in any strict sense from simpler systems, but he does argue that the characteristic properties of human language can be found in animal communication systems, although "we cannot at the moment say definitely that they are all present in one particular animal." The characteristics shared by human and animal language are the properties of being "purposive," "syntactic," and "propositional." Language is purposive "in that there is nearly always in human speech a definite intention of getting something over to somebody else, altering his behavior, his thoughts, or his general attitude toward a situation." Human language is "syntactic" in that an utterance is a performance with an internal organization, with structure and coherence. It is "propositional" in that it transmits information. In this sense, then, both human language and animal communication are purposive, syntactic, and propositional.

All this may be true, but it establishes very little, since when we move to the level of abstraction at which human language and animal communication fall together, almost all other behavior is included as well. Consider walking: clearly, walking is purposive behavior, in the most general sense of "purposive." Walking is also "syntactic" in the sense just defined, as, in fact, Karl Lashley pointed out a long time ago in his important discussion of serial order in behavior, to which I referred in the first lecture.[4] Furthermore, it can certainly be informative; for example, I can signal my interest in reaching a certain goal by the speed or intensity with which I walk.

It is, incidentally, precisely in this manner that the examples of animal communication that Thorpe presents are "propositional." He cites as an example the song of the European robin, in which the rate of alternation of high and low pitch signals the intention of the bird to defend its territory; the higher the rate of alternation, the greater the intention to defend the territory. The example is interesting, but it seems to me to show very clearly the hopelessness of the attempt to relate human language to animal communication. Every animal communication system that is known (if we disregard some science fiction about dolphins) uses one of two basic principles: either it consists of a fixed, finite number of signals, each associated with a specific range of behavior or emotional state, as is illustrated in the extensive primate studies that have been carried out by Japanese scientists for the past several years; or it makes use of a fixed, finite number of linguistic dimensions, each of which is associated with a particular nonlinguistic dimension in such a way that selection of a point along the linguistic dimension determines and signals a certain point along the associated nonlinguistic dimension. The latter is the principle realized in Thorpe's

[4] K. S. Lashley, "The Problem of Serial Order in Behavior," in L. A. Jeffress, ed., *Cerebral Mechanisms in Behavior* (New York: Wiley, 1951), pp. 112–36.

bird-song example. Rate of alternation of high and low pitch is a linguistic dimension correlated with the nonlinguistic dimension of intention to defend a territory. The bird signals its intention to defend a territory by selecting a correlated point along the linguistic dimension of pitch alternation – I use the word "select" loosely, of course. The linguistic dimension is abstract, but the principle is clear. A communication system of the second type has an indefinitely large range of potential signals, as does human language. The mechanism and principle, however, are entirely different from those employed by human language to express indefinitely many new thoughts, intentions, feelings, and so on. It is not correct to speak of a "deficiency" of the animal system, in terms of range of potential signals; rather the opposite, since the animal system admits in principle of continuous variation along the linguistic dimension (insofar as it makes sense to speak of "continuity" in such a case), whereas human language is discrete. Hence, the issue is not one of "more" or "less," but rather of an entirely different principle of organization. When I make some arbitrary statement in a human language – say, that "the rise of supranational corporations poses new dangers for human freedom" – I am not selecting a point along some linguistic dimension that signals a corresponding point along an associated nonlinguistic dimension, nor am I selecting a signal from a finite behavioral repertoire, innate or learned.

Furthermore, it is wrong to think of human use of language as characteristically informative, in fact or in intention. Human language can be used to inform or mislead, to clarify one's own thoughts or to display one's cleverness, or simply for play. If I speak with no concern for modifying your behavior or thoughts, I am not using language any less than if I say exactly the same things *with* such intention. If we hope to understand human language and the psychological capacities on which it rests, we must first ask what it is, not how or for what purposes it is used. When we ask what human language is, we find no striking similarity to animal communication systems. There is nothing useful to be said about behavior or thought at the level of abstraction at which animal and human communication fall together. The examples of animal communication that have been examined to date do share many of the properties of human gestural systems, and it might be reasonable to explore the possibility of direct connection in this case. But human language, it appears, is based on entirely different principles. This, I think, is an important point, often overlooked by those who approach human language as a natural, biological phenomenon; in particular, it seems rather pointless, for these reasons, to speculate about the evolution of human language from simpler systems – perhaps as absurd as it would be to speculate about the "evolution" of atoms from clouds of elementary particles.

As far as we know, possession of human language is associated with a specific type of mental organization, not simply a higher degree of intelligence.

There seems to be no substance to the view that human language is simply a more complex instance of something to be found elsewhere in the animal world. This poses a problem for the biologist, since, if true, it is an example of true "emergence" – the appearance of a qualitatively different phenomenon at a specific stage of complexity of organization. Recognition of this fact, though formulated in entirely different terms, is what motivated much of the classical study of language by those whose primary concern was the nature of mind. And it seems to me that today there is no better or more promising way to explore the essential and distinctive properties of human intelligence than through the detailed investigation of the structure of this unique human possession. A reasonable guess, then, is that if empirically adequate generative grammars can be constructed and the universal principles that govern their structure and organization determined, then this will be an important contribution to human psychology, in ways to which I will turn directly, in detail.

In the course of these lectures I have mentioned some of the classical ideas regarding language structure and contemporary efforts to deepen and extend them. It seems clear that we must regard linguistic competence – knowledge of a language – as an abstract system underlying behavior, a system constituted by rules that interact to determine the form and intrinsic meaning of a potentially infinite number of sentences. Such a system – a generative grammar – provides an explication of the Humboldtian idea of "form of language," which in an obscure but suggestive remark in his great posthumous work, *Über die Verschiedenheit des Menschlichen Sprachbaues*, Humboldt defines as "that constant and unvarying system of processes underlying the mental act of raising articulated structurally organized signals to an expression of thought." Such a grammar defines a language in the Humboldtian sense, namely as "a recursively generated system, where the laws of generation are fixed and invariant, but the scope and the specific manner in which they are applied remain entirely unspecified."

In each such grammar there are particular, idiosyncratic elements, selection of which determines one specific human language; and there are general universal elements, conditions on the form and organization of any human language, that form the subject matter for the study of "universal grammar." Among the principles of universal grammar are those I discussed in the preceding lecture – for example, the principles that distinguish deep and surface structure and that constrain the class of transformational operations that relate them. Notice, incidentally, that the existence of definite principles of universal grammar makes possible the rise of the new field of mathematical linguistics, a field that submits to abstract study the class of generative systems meeting the conditions set forth in universal grammar. This inquiry aims to elaborate the formal properties of any possible human language. The field is in its infancy; it is only in the last decade that the possibility of such an enterprise has been envisioned. It

has some promising initial results, and it suggests one possible direction for future research that might prove to be of great importance. Thus, mathematical linguistics seems for the moment to be in a uniquely favorable position, among mathematical approaches in the social and psychological sciences, to develop not simply as a theory of data, but as the study of highly abstract principles and structures that determine the character of human mental processes. In this case, the mental processes in question are those involved in the organization of one specific domain of human knowledge, namely knowledge of language.

The theory of generative grammar, both particular and universal, points to a conceptual lacuna in psychological theory that I believe is worth mentioning. Psychology conceived as "behavioral science" has been concerned with behavior and acquisition or control of behavior. It has no concept corresponding to "competence," in the sense in which competence is characterized by a generative grammar. The theory of learning has limited itself to a narrow and surely inadequate concept of what is learned – namely a system of stimulus-response connections, a network of associations, a repertoire of behavioral items, a habit hierarchy, or a system of dispositions to respond in a particular way under specifiable stimulus conditions.[5] Insofar as behavioral psychology has been applied to education or therapy, it has correspondingly limited itself to this concept of "what is learned." But a generative grammar cannot be characterized in these terms. What is necessary, in addition to the concept of behavior and learning, is a concept of what is learned – a notion of competence – that lies beyond the conceptual limits of behaviorist psychological theory. Like much of modern linguistics and modern philosophy of language, behaviorist psychology has quite consciously accepted methodological restrictions that do not permit the study of systems of the necessary complexity and abstractness.[6] One important future

[5] This limitation is revealed, for example, in such statements as this from W. M. Wiest, in "Recent Criticisms of Behaviorism and Learning," in *Psychological Bulletin*, Vol. 67, No. 3, 1967, pp. 214–25: "An empirical demonstration . . . that a child has learned the rules of grammar would be his exhibiting the verbal performance called 'uttering the rules of grammar.' That this performance is not usually acquired without special training is attested to by many grammar school teachers. One may even speak quite grammatically without having literally learned the rules of grammar." Wiest's inability to conceive of another sense in which the child may be said to have learned the rules of grammar testifies to the conceptual gap we are discussing. Since he refuses to consider the question of *what* is learned, and to clarify this notion before asking *how* it is learned, he can only conceive of "grammar" as the "behavioral regularities in the understanding and production of speech" – a characterization that is perfectly empty, as it stands, there being no "behavioral regularities" associated with (let alone "in") the understanding and production of speech. One cannot quarrel with the desire of some investigators to study "the acquisition and maintenance of *actual occurrences of verbal behavior*" (*ibid.*). It remains to be demonstrated that this study has something to do with the study of language. As of now, I see no indication that this claim can be substantiated.

[6] See my paper, "Some Empirical Assumptions in Modern Philosophy of Language," in S. Morgenbesser, P. Suppes, and M. White, eds., *Essays in Honor of Ernest Nagel* (New York: St. Martin's, 1969), for a discussion of the work of Quine and Wittgenstein from this point of view.

contribution of the study of language to general psychology may be to focus attention on this conceptual gap and to demonstrate how it may be filled by the elaboration of a system of underlying competence in one domain of human intelligence.

There is an obvious sense in which any aspect of psychology is based ultimately on the observation of behavior. But it is not at all obvious that the study of learning should proceed directly to the investigation of factors that control behavior or of conditions under which a "behavioral repertoire" is established. It is first necessary to determine the significant characteristics of this behavioral repertoire, the principles on which it is organized. A meaningful study of learning can proceed only after this preliminary task has been carried out and has led to a reasonably well-confirmed theory of underlying competence – in the case of language, to the formulation of the generative grammar that underlies the observed use of language. Such a study will concern itself with the relation between the data available to the organism and the competence that it acquires; only to the extent that the abstraction to competence has been successful – in the case of language, to the extent that the postulated grammar is "descriptively adequate" in the sense described in Lecture 2 – can the investigation of learning hope to achieve meaningful results. If, in some domain, the organization of the behavioral repertoire is quite trivial and elementary, then there will be little harm in avoiding the intermediate stage of theory construction, in which we attempt to characterize accurately the competence that is acquired. But one cannot count on this being the case, and in the study of language it surely is not the case. With a richer and more adequate characterization of "what is learned" – of the underlying competence that constitutes the "final state" of the organism being studied – it may be possible to approach the task of constructing a theory of learning that will be much less restricted in scope than modern behavioral psychology has proved to be. Surely it is pointless to accept methodological strictures that preclude such an approach to problems of learning.

Are there other areas of human competence where one might hope to develop a fruitful theory, analogous to generative grammar? Although this is a very important question, there is very little that can be said about it today. One might, for example, consider the problem of how a person comes to acquire a certain concept of three-dimensional space, or an implicit "theory of human action," in similar terms. Such a study would begin with the attempt to characterize the implicit theory that underlies actual performance and would then turn to the question of how this theory develops under the given conditions of time and access to data – that is, in what way the resulting system of beliefs is determined by the interplay of available data, "heuristic procedures," and the innate schematism that restricts and conditions the form of the acquired system. At the moment, this is nothing more than a sketch of a program of research.

There have been some attempts to study the structure of other, language-like systems – the study of kinship systems and folk taxonomies comes to mind, for example. But so far, at least, nothing has been discovered that is even roughly comparable to language in these domains. No one, to my knowledge, has devoted more thought to this problem than Lévi-Strauss. For example, his recent book on the categories of primitive mentality[7] is a serious and thoughtful attempt to come to grips with this problem. Nevertheless, I do not see what conclusions can be reached from a study of his materials beyond the fact that the savage mind attempts to impose some organization on the physical world – that humans classify, if they perform any mental acts at all. Specifically, Lévi-Strauss's well-known critique of totemism seems to reduce to little more than this conclusion.

Lévi-Strauss models his investigations quite consciously on structural linguistics, particularly on the work of Troubetzkoy and Jakobson. He repeatedly and quite correctly emphasizes that one cannot simply apply procedures analogous to those of phonemic analysis to subsystems of society and culture. Rather, he is concerned with structures "where they may be found . . . in the kinship system, political ideology, mythology, ritual, art," and so on,[8] and he wishes to examine the formal properties of these structures in their own terms. But several reservations are necessary when structural linguistics is used as a model in this way. For one thing, the structure of a phonological system is of very little interest as a formal object; there is nothing of significance to be said, from a formal point of view, about a set of forty-odd elements cross-classified in terms of eight or ten features. The significance of structuralist phonology, as developed by Troubetzkoy, Jakobson, and others, lies not in the formal properties of phonemic systems but in the fact that a fairly small number of features that can be specified in absolute, language-independent terms appear to provide the basis for the organization of all phonological systems. The achievement of structuralist phonology was to show that the phonological rules of a great variety of languages apply to classes of elements that can be simply characterized in terms of these features; that historical change affects such classes in a uniform way; and that the organization of features plays a basic role in the use and acquisition of language. This was a discovery of the greatest importance, and it provides the groundwork for much of contemporary linguistics. But if we abstract away from the specific universal set of features and the rule systems in which they function, little of any significance remains.

Furthermore, to a greater and greater extent, current work in phonology is demonstrating that the real richness of phonological systems lies not in the structural patterns of phonemes but rather in the intricate systems of rules by

[7] C. Lévi-Strauss, *The Savage Mind* (Chicago: University of Chicago Press, 1967).
[8] C. Lévi-Strauss, *Structural Anthropology* (New York: Basic Books, 1963), p. 85.

which these patterns are formed, modified, and elaborated.[9] The structural patterns that arise at various stages of derivation are a kind of epiphenomenon. The system of phonological rules makes use of the universal features in a fundamental way,[10] but it is the properties of the systems of rules, it seems to me, that really shed light on the specific nature of the organization of language. For example, there appear to be very general conditions, such as the principle of cyclic ordering (discussed in the preceding lecture) and others that are still more abstract, that govern the application of these rules, and there are many interesting and unsolved questions as to how the choice of rules is determined by intrinsic, universal relations among features. Furthermore, the idea of a mathematical investigation of language structures, to which Lévi-Strauss occasionally alludes, becomes meaningful only when one considers systems of rules with infinite generative capacity. There is nothing to be said about the abstract structure of the various patterns that appear at various stages of derivation. If this is correct, then one cannot expect structuralist phonology, in itself, to provide a useful model for investigation of other cultural and social systems.

In general, the problem of extending concepts of linguistic structure to other cognitive systems seems to me, for the moment, in not too promising a state, although it is no doubt too early for pessimism.

Before turning to the general implications of the study of linguistic competence and, more specifically, to the conclusions of universal grammar, it is well to make sure of the status of these conclusions in the light of current knowledge of the possible diversity of language. In my first lecture, I quoted the remarks of William Dwight Whitney about what he referred to as "the infinite diversity of human speech," the boundless variety that, he maintained, undermines the claims of philosophical grammar to psychological relevance.

Philosophical grammarians had typically maintained that languages vary little in their deep structures, though there may be wide variability in surface manifestations. Thus there is, in this view, an underlying structure of grammatical relations and categories, and certain aspects of human thought and mentality are essentially invariant across languages, although languages may differ as to whether they express the grammatical relations formally by inflection or word order, for example. Furthermore, an investigation of their work indicates that the underlying recursive principles that generate deep structure were assumed to be restricted in certain ways – for example, by the condition that new structures are formed only by the insertion of new "propositional content," new structures that themselves correspond to actual simple sentences, in fixed positions in already

[9] See discussion in the preceding lecture and the references cited there.

[10] The study of universal features is itself in considerable flux. See N. Chomsky and M. Halle, *The Sound Pattern of English* (New York: Harper & Row, 1968), Chapter 7, for recent discussion.

formed structures. Similarly, the grammatical transformations that form surface structures through reordering, ellipsis, and other formal operations must themselves meet certain fixed general conditions, such as those discussed in the preceding lecture. In short, the theories of philosophical grammar, and the more recent elaborations of these theories, make the assumption that languages will differ very little, despite considerable diversity in superficial realization, when we discover their deeper structures and unearth their fundamental mechanisms and principles.

It is interesting to observe that this assumption persisted even through the period of German romanticism, which was, of course, much preoccupied with the diversity of cultures and with the many rich possibilities for human intellectual development. Thus, Wilhelm von Humboldt, who is now best remembered for his ideas concerning the variety of languages and the association of diverse language structures with divergent "world-views," nevertheless held firmly that underlying any human language we will find a system that is universal, that simply expresses man's unique intellectual attributes. For this reason, it was possible for him to maintain the rationalist view that language is not really learned – certainly not taught – but rather develops "from within," in an essentially predetermined way, when the appropriate environmental conditions exist. One cannot really teach a first language, he argued, but can only "provide the thread along which it will develop of its own accord," by processes more like maturation than learning. This Platonistic element in Humboldt's thought is a pervasive one; for Humboldt, it was as natural to propose an essentially Platonistic theory of "learning" as it was for Rousseau to found his critique of repressive social institutions on a conception of human freedom that derives from strictly Cartesian assumptions regarding the limitations of mechanical explanation. And in general it seems appropriate to construe both the psychology and the linguistics of the romantic period as in large part a natural outgrowth of rationalist conceptions.[11]

The issue raised by Whitney against Humboldt and philosophical grammar in general is of great significance with respect to the implications of linguistics for general human psychology. Evidently, these implications can be truly far-reaching only if the rationalist view is essentially correct, in which case the structure of language can truly serve as a "mirror of mind," in both its particular and its universal aspects. It is widely believed that modern anthropology has established the falsity of the assumptions of the rationalist universal grammarians by demonstrating through empirical study that languages may, in fact, exhibit the widest diversity. Whitney's claims regarding the diversity of languages are reiterated throughout the modern period; Martin Joos, for example,

[11] For some discussion of these matters, see my *Cartesian Linguistics* (New York: Harper & Row, 1966).

is simply expressing the conventional wisdom when he takes the basic conclusion of modern anthropological linguistics to be that "languages can differ without limit as to either extent or direction."[12]

The belief that anthropological linguistics has demolished the assumptions of universal grammar seems to me to be quite false in two important respects. First, it misinterprets the views of classical rationalist grammar, which held that languages are similar only at the deeper level, the level at which grammatical relations are expressed and at which the processes that provide for the creative aspect of language use are to be found. Second, this belief seriously misinterprets the findings of anthropological linguistics, which has, in fact, restricted itself almost completely to fairly superficial aspects of language structure.

To say this is not to criticize anthropological linguistics, a field that is faced with compelling problems of its own – in particular, the problem of obtaining at least some record of the rapidly vanishing languages of the primitive world. Nevertheless, it is important to bear in mind this fundamental limitation on its achievements in considering the light it can shed on the theses of universal grammar. Anthropological studies (like structural linguistic studies in general) do not attempt to reveal the underlying core of generative processes in language – that is, the processes that determine the deeper levels of structure and that constitute the systematic means for creating ever novel sentence types. Therefore, they obviously cannot have any real bearing on the classical assumption that these underlying generative processes vary only slightly from language to language. In fact, what evidence is now available suggests that if universal grammar has serious defects, as indeed it does from a modern point of view, then these defects lie in the failure to recognize the abstract nature of linguistic structure and to impose sufficiently strong and restrictive conditions on the form of any human language. And a characteristic feature of current work in linguistics is its concern for linguistic universals of a sort that can only be detected through a detailed investigation of particular languages, universals governing properties of language that are simply not accessible to investigation within the restricted framework that has been adopted, often for very good reasons, within anthropological linguistics.

I think that if we contemplate the classical problem of psychology, that of accounting for human knowledge, we cannot avoid being struck by the enormous disparity between knowledge and experience – in the case of language, between the generative grammar that expresses the linguistic competence of the

[12] M. Joos, ed., *Readings in Linguistics*, 4th edn. (Chicago: University of Chicago Press, 1966), p. 228. This is put forth as the "Boas Tradition." American linguistics, Joos maintains, "got its decisive direction when it was decided that an indigenous language could be described without any preexistent scheme of what a language must be . . ." (p. 1). Of course this could not literally be true – the procedures of analysis themselves express a hypothesis concerning the possible diversity of language. But there is, nevertheless, much justice in Joos's characterization.

native speaker and the meager and degenerate data on the basis of which he has constructed this grammar for himself. In principle the theory of learning should deal with this problem; but in fact it bypasses the problem, because of the conceptual gap that I mentioned earlier. The problem cannot even be formulated in any sensible way until we develop the concept of competence, alongside the concepts of learning and behavior, and apply this concept in some domain. The fact is that this concept has so far been extensively developed and applied only in the study of human language. It is only in this domain that we have at least the first steps toward an account of competence, namely the fragmentary generative grammars that have been constructed for particular languages. As the study of language progresses, we can expect with some confidence that these grammars will be extended in scope and depth, although it will hardly come as a surprise if the first proposals are found to be mistaken in fundamental ways.

Insofar as we have a tentative first approximation to a generative grammar for some language, we can for the first time formulate in a useful way the problem of origin of knowledge. In other words, we can ask the question, What initial structure must be attributed to the mind that enables it to construct such a grammar from the data of sense? Some of the empirical conditions that must be met by any such assumption about innate structure are moderately clear. Thus, it appears to be a species-specific capacity that is essentially independent of intelligence, and we can make a fairly good estimate of the amount of data that is necessary for the task to be successfully accomplished. We know that the grammars that are in fact constructed vary only slightly among speakers of the same language, despite wide variations not only in intelligence but also in the conditions under which language is acquired. As participants in a certain culture, we are naturally aware of the great differences in ability to use language, in knowledge of vocabulary, and so on that result from differences in native ability and from differences in conditions of acquisition; we naturally pay much less attention to the similarities and to common knowledge, which we take for granted. But if we manage to establish the requisite psychic distance, if we actually compare the generative grammars that must be postulated for different speakers of the same language, we find that the similarities that we take for granted are quite marked and that the divergences are few and marginal. What is more, it seems that dialects that are superficially quite remote, even barely intelligible on first contact, share a vast central core of common rules and processes and differ very slightly in underlying structures, which seem to remain invariant through long historical eras. Furthermore, we discover a substantial system of principles that do not vary among languages that are, as far as we know, entirely unrelated.

The central problems in this domain are empirical ones that are, in principle at least, quite straightforward, difficult as they may be to solve in a satisfactory way. We must postulate an innate structure that is rich enough to account for

the disparity between experience and knowledge, one that can account for the construction of the empirically justified generative grammars within the given limitations of time and access to data. At the same time, this postulated innate mental structure must not be so rich and restrictive as to exclude certain known languages. There is, in other words, an upper bound and a lower bound on the degree and exact character of the complexity that can be postulated as innate mental structure. The factual situation is obscure enough to leave room for much difference of opinion over the true nature of this innate mental structure that makes acquisition of language possible. However, there seems to me to be no doubt that this is an empirical issue, one that can be resolved by proceeding along the lines that I have just roughly outlined.

My own estimate of the situation is that the real problem for tomorrow is that of discovering an assumption regarding innate structure that is sufficiently rich, not that of finding one that is simple or elementary enough to be "plausible." There is, as far as I can see, no reasonable notion of "plausibility," no a priori insight into what innate structures are permissible, that can guide the search for a "sufficiently elementary assumption." It would be mere dogmatism to maintain without argument or evidence that the mind is simpler in its innate structure than other biological systems, just as it would be mere dogmatism to insist that the mind's organization must necessarily follow certain set principles, determined in advance of investigation and maintained in defiance of any empirical findings. I think that the study of problems of mind has been very definitely hampered by a kind of apriorism with which these problems are generally approached. In particular, the empiricist assumptions that have dominated the study of acquisition of knowledge for many years seem to me to have been adopted quite without warrant and to have no special status among the many possibilities that one might imagine as to how the mind functions.

In this connection, it is illuminating to follow the debate that has arisen since the views I have just sketched were advanced a few years ago as a program of research – I should say, since this position was resurrected, because to a significant extent it is the traditional rationalist approach, now amplified and sharpened and made far more explicit in terms of the tentative conclusions that have been reached in the recent study of linguistic competence. Two outstanding American philosophers, Nelson Goodman and Hilary Putnam, have made recent contributions to this discussion – both misconceived, in my opinion, but instructive in the misconceptions that they reveal.[13]

[13] N. Goodman, "The Epistemological Argument," and H. Putnam, "The Innateness Hypothesis and Explanatory Models in Linguistics." Together with a paper of mine, these were presented at the Innate Ideas Symposium of the American Philosophical Association and the Boston Colloquium for the Philosophy of Science in December 1966. The three essays appear in *Synthèse*, Vol. 17, No. 1, 1967, pp. 2–28, and in R. S. Cohen and W. M. Wartofsky, eds., *Boston Studies in the Philosophy of Science*, Vol. III (New York: Humanities, 1968),

Goodman's treatment of the question suffers first from an historical misunderstanding and second from a failure to formulate correctly the exact nature of the problem of acquisition of knowledge. His historical misunderstanding has to do with the issue between Locke and whomever Locke thought he was criticizing in his discussion of innate ideas. According to Goodman, "Locke made . . . acutely clear" that the doctrine of innate ideas is "false or meaningless." In fact, however, Locke's critique had little relevance to any familiar doctrine of the seventeenth century. The arguments that Locke gave were considered and dealt with in quite a satisfactory way in the earliest seventeenth-century discussions of innate ideas, for example those of Lord Herbert and Descartes, both of whom took for granted that the system of innate ideas and principles would not function unless appropriate stimulation took place. For this reason, Locke's arguments, none of which took cognizance of this condition, are without force;[14] for some reason, he avoided the issues that had been discussed in the preceding half-century. Furthermore, as Leibnitz observed, Locke's willingness to make use of a principle of "reflection" makes it almost impossible to distinguish his approach from that of the rationalists, except for his failure to take even those steps suggested by his predecessors toward specifying the character of this principle.

But, historical issues aside, I think that Goodman misconstrues the substantive problem as well. He argues that first-language learning poses no real problem, because prior to first-language learning the child has already acquired the rudiments of a symbolic system in his ordinary dealings with the environment. Hence, first-language learning is analogous to second-language learning in that the fundamental step has already been taken, and details can be elaborated within an already existing framework. This argument might have some force if it were possible to show that the specific properties of grammar – say, the distinction of deep and surface structure, the specific properties of grammatical transformations, the principles of rule ordering, and so on – were present in some form in these already acquired prelinguistic "symbolic systems." But since there is not the slightest reason to believe that this is so, the argument

pp. 81–107. A more extensive discussion of the papers of Putnam and Goodman, along with a number of others, appears in my contribution to the symposium "Linguistics and Philosophy," New York University, April 1968, in S. Hook, ed., *Philosophy and Language* (New York: New York University Press, 1969). The essay is reprinted in this volume.

[14] This observation is a commonplace. See, for example, the commentary by A. C. Fraser in his edition of Locke's *Essay Concerning Human Understanding*, 1894 (reprinted by Dover, 1959), notes 1 and 2, Chapter 1 (p. 38 of the Dover edition). As Fraser notes, Descartes' position is one "which Locke's argument always fails to reach . . . Locke assails [the hypothesis of innate ideas] . . . in its crudest form, in which it is countenanced by no eminent advocate." Goodman is free to use the term "innate idea" in conformity with Locke's misinterpretation of the doctrine if he wishes, but not to charge "sophistry," as he does, when others examine and develop rationalist doctrine in the form in which it was actually presented.

collapses. It is based on an equivocation similar to that discussed earlier in connection with the argument that language evolved from animal communication. In that case, as we observed, the argument turned on a metaphorical use of the term "language." In Goodman's case, the argument is based entirely on a vague use of the term "symbolic system," and it collapses as soon as we attempt to give this term a precise meaning. If it were possible to show that these prelinguistic symbolic systems share certain significant properties with natural language, we could then argue that these properties of natural language are acquired by analogy. Of course, we would then face the problem of explaining how the prelinguistic symbolic systems developed these properties. But since no one has succeeded in showing that the fundamental properties of natural language – those discussed in Lecture 2, for example – appear in prelinguistic symbolic systems or any others, the latter problem does not arise.

According to Goodman, the reason why the problem of second-language learning is different from that of first-language learning is that "once one language is available," it "can be used for giving explanation and instruction." He then goes on to argue that "acquisition of an initial language is acquisition of a secondary symbolic system" and is quite on a par with normal second-language acquisition. The primary symbolic systems to which he refers are "rudimentary prelinguistic symbolic systems in which gestures and sensory and perceptual occurrences of all sorts function as signs." But evidently these prelinguistic symbolic systems cannot be "used for giving explanation and instruction" in the way a first language can be used in second-language instruction. Therefore, even on his own grounds, Goodman's argument is incoherent.

Goodman maintains that "the claim we are discussing cannot be experimentally tested even when we have an acknowledged example of a 'bad' language" and that "the claim has not even been formulated to the extent of citation of a single general property of 'bad' languages." The first of these conclusions is correct, in his sense of "experimental test," namely a test in which we "take an infant at birth, isolate it from all the influences of our language-bound culture, and attempt to inculcate it with one of the 'bad' artificial languages." Obviously this is not feasible. But there is no reason why we should be dismayed by the impossibility of carrying out such a test as this. There are many other ways – for example, those discussed in Lecture 2 and the references cited there – in which evidence can be obtained concerning the properties of grammars and conclusions regarding the general properties of such grammars can be put to empirical test. Any such conclusion immediately specifies, correctly or incorrectly, certain properties of "bad" languages. Since there are dozens of papers and books that attempt to formulate such properties, his second claim, that not "a single general property of 'bad' languages" has been formulated, is rather surprising. One might try to show that these attempts are misguided or questionable, but one can hardly maintain seriously that they do not exist. Any formulation of

a principle of universal grammar makes a strong empirical claim, which can be falsified by finding counter-instances in some human language, along the lines of the discussion in Lecture 2. In linguistics, as in any other field, it is only in such indirect ways as this that one can hope to find evidence bearing on nontrivial hypotheses. Direct experimental tests of the sort that Goodman mentions are rarely possible, a matter that may be unfortunate but is nevertheless characteristic of most research.

At one point Goodman remarks, correctly, that even though "for certain remarkable facts I have no alternative explanation . . . that alone does not dictate acceptance of whatever theory may be offered; for the theory might be worse than none. Inability to explain a fact does not condemn me to accept an intrinsically repugnant and incomprehensible theory." But now consider the theory of innate ideas that Goodman regards as "intrinsically repugnant and incomprehensible." Notice, first, that the theory is obviously not "incomprehensible," on his terms. Thus he appears to be willing, in this article, to accept the view that in some sense the mature mind contains ideas; it is obviously not "incomprehensible," then, that some of these ideas are "implanted in the mind as original equipment," to use his phraseology. And if we turn to the actual doctrine as developed in rationalist philosophy, rather than Locke's caricature, the theory becomes even more obviously comprehensible. There is nothing incomprehensible in the view that stimulation provides the occasion for the mind to apply certain innate interpretive principles, certain concepts that proceed from "the power of understanding" itself, from the faculty of thinking rather than from external objects directly. To take an example from Descartes (*Reply to Objections*, V):

> When first in infancy we see a triangular figure depicted on paper, this figure cannot show us how a real triangle ought to be conceived, in the way in which geometricians consider it, because the true triangle is contained in this figure, just as the statue of Mercury is contained in a rough block of wood. But because we already possess within us the idea of a true triangle, and it can be more easily conceived by our mind than the more complex figure of the triangle drawn on paper, we, therefore, when we see the composite figure, apprehend not it itself, but rather the authentic triangle.[15]

In this sense the idea of a triangle is innate. Surely the notion is comprehensible; there would be no difficulty, for example, in programing a computer to react to stimuli along these lines (though this would not satisfy Descartes, for other reasons). Similarly, there is no difficulty in principle in programing a computer with a schematism that sharply restricts the form of a generative grammar, with an evaluation procedure for grammars of the given form, with a

[15] E. S. Haldane and G. R. T. Ross, eds., *Descartes' Philosophical Works*, 1911 (reprinted by Dover, 1955). The citation, and the preceding remarks, appear in my contribution to the Innate Ideas Symposium of December 1966 (see note 13).

technique for determining whether given data are compatible with a grammar of the given form, with a fixed substructure of entities (such as distinctive features), rules, and principles, and so on – in short, with a universal grammar of the sort that has been proposed in recent years. For reasons that I have already mentioned, I believe that these proposals can be properly regarded as a further development of classical rationalist doctrine, as an elaboration of some of its main ideas regarding language and mind. Of course, such a theory will be "repugnant" to one who accepts empiricist doctrine and regards it as immune to question or challenge. It seems to me that this is the heart of the matter.

Putnam's paper (see note 13) deals more directly with the points at issue, but it seems to me that his arguments are also inconclusive, because of certain incorrect assumptions that he makes about the nature of the acquired grammars. Putnam assumes that on the level of phonetics the only property proposed in universal grammar is that a language has "a short list of phonemes." This, he argues, is not a similarity among languages that requires elaborate explanatory hypotheses. The conclusion is correct; the assumption is quite wrong. In fact, as I have now pointed out several times, very strong empirical hypotheses have been proposed regarding the specific choice of universal features, conditions on the form and organization of phonological rules, conditions on rule application, and so on. If these proposals are correct or near correct, then "similarities among languages" at the level of sound structure are indeed remarkable and cannot be accounted for simply by assumptions about memory capacity, as Putnam suggests.

Above the level of sound structure, Putnam assumes that the only significant properties of language are that they have proper names, that the grammar contains a phrase structure component, and that there are rules "abbreviating" sentences generated by the phrase structure component. He argues that the nature of the phrase structure component is determined by the existence of proper names; that the existence of a phrase structure component is explained by the fact that "all the natural measures of complexity of an algorithm – size of the machine table, length of computations, time, and space required for the computation – lead to the . . . result"; that phrase structure systems provide the "algorithms which are 'simplest' for virtually any computing system," hence also "for naturally evolved 'computing systems'"; and that there is nothing surprising in the fact that languages contain rules of abbreviation.

Each of the three conclusions involves a false assumption. From the fact that a phrase structure system contains proper names one can conclude almost nothing about its other categories. In fact, there is much dispute at the moment about the general properties of the underlying phrase structure system for natural languages; the dispute is not in the least resolved by the existence of proper names.

As to the second point, it is simply untrue that all measures of complexity and speed of computation lead to phrase structure rules as the "simplest possible algorithm." The only existing results that are even indirectly relevant show that context-free phrase structure grammars (a reasonable model for rules generating deep structures, when we exclude the lexical items and the distributional conditions they meet) receive an automata-theoretic interpretation as nondeterministic pushdown storage automata, but the latter is hardly a "natural" notion from the point of view of "simplicity of algorithms" and so forth. In fact, it can be argued that the somewhat similar but not formally related concept of real-time deterministic automation is far more "natural" in terms of time and space conditions on computation.[16]

However, it is pointless to pursue this topic, because what is at stake is not the "simplicity" of phrase structure grammars but rather of transformational grammars with a phrase structure component that plays a role in generating deep structures. And there is absolutely no mathematical concept of "ease of computation" or "simplicity of algorithm" that even vaguely suggests that such systems may have some advantage over the kinds of automata that have been seriously investigated from this point of view – for example, finite state automata, linear bounded automata, and so on. The basic concept of "structure-dependent operation" has never even been considered in a strictly mathematical concept. The source of this confusion is a misconception on Putnam's part as to the nature of grammatical transformations. They are not rules that "abbreviate" sentences; rather, they are operations that form surface structures from underlying deep structures, in such ways as are illustrated in the preceding lecture and the references there cited.[17] Hence, to show that transformational grammars are the "simplest possible," one would have to demonstrate that the "optimal" computing system would take a string of symbols as input and determine its surface structure, its underlying deep structure, and the sequence of transformational operations that relates them. Nothing of the sort has been shown; in fact, the question has never even been raised.

Putnam argues that even if significant uniformities among languages were to be discovered, there would be a simpler explanation than the hypothesis of an innate universal grammar, namely their common origin. But this proposal

[16] For some discussion of these matters, see my "Formal Properties of Grammars," in R. D. Luce, R. Bush, and E. Galanter, eds., *Handbook of Mathematical Psychology*, Vol. II (New York: Wiley, 1963). For a more extensive discussion of the automata-theoretic framework, see R. J. Nelson, *Introduction to Automata* (New York: Wiley, 1968). A detailed presentation of properties of context-free grammars is given in S. Ginsburg, *The Mathematical Theory of Context-Free Languages* (New York: McGraw-Hill, 1966). There have been a number of studies of speed of computation, simplicity of algorithms, and so on, but none of them has any bearing on the issue under discussion.

[17] See note 10 of Lecture 2, p. 29, for further comment.

involves a serious misunderstanding of the problem at issue. The grammar of a language must be discovered by the child from the data presented to him. As noted earlier, the empirical problem is to find a hypothesis about initial structure rich enough to account for the fact that a specific grammar is constructed by the child, but not so rich as to be falsified by the known diversity of language. Questions of common origin are of potential relevance to this empirical issue in only one respect: if the existing languages are not a "fair sample" of the "possible languages," we may be led mistakenly to propose too narrow a schema for universal grammar. However, as I mentioned earlier, the empirical problem that we face today is that no one has been able to devise an initial hypothesis rich enough to account for the acquisition by the child of the grammar that we are, apparently, led to attribute to him when we try to account for his ability to use the language in the normal way. The assumption of common origin contributes nothing to explaining how this achievement is possible. In short, the language is "reinvented" each time it is learned, and the empirical problem to be faced by the theory of learning is how this invention of grammar can take place.

Putnam does face this problem and suggests that there might be "general multipurpose learning strategies" that account for this achievement. It is, of course, an empirical question whether the properties of the "language faculty" are specific to language or are merely a particular case of much more general mental faculties (or learning strategies). This is a problem that has been discussed earlier in this lecture, inconclusively and in a slightly different context. Putnam takes for granted that it is only general "learning strategies" that are innate but suggests no grounds for this empirical assumption. As I have argued earlier, a nondogmatic approach to this problem can be pursued, without reliance on unargued assumptions of this sort – that is, through the investigation of specific areas of human competence, such as language, followed by the attempt to devise a hypothesis that will account for the development of this competence. If we discover through such investigation that the same "learning strategies" are sufficient to account for the development of competence in various domains, we will have reason to believe that Putnam's assumption is correct. If we discover that the postulated innate structures differ from case to case, the only rational conclusion would be that a model of mind must involve separate "faculties," with unique or partially unique properties. I cannot see how anyone can resolutely insist on one or the other conclusion in the light of the evidence now available to us. But one thing is quite clear: Putnam has no justification for his final conclusion, that "invoking 'Innateness' only postpones the problem of learning; it does not solve it." Invoking an innate representation of universal grammar does solve the problem of learning, if it is true that this is the basis for language acquisition, as it well may be. If, on the other hand, there are general learning strategies that account for the acquisition of grammatical knowledge, then postulation of an innate universal grammar will not "postpone" the problem

of learning, but will rather offer an incorrect solution to this problem. The issue is an empirical one of truth or falsity, not a methodological one of states of investigation.[18]

To summarize, it seems to me that neither Goodman nor Putnam offers a serious counterargument to the proposals concerning innate mental structure that have been advanced (tentatively, of course, as befits empirical hypotheses) or suggests a plausible alternative approach, with empirical content, to the problem of acquisition of knowledge.

Assuming the rough accuracy of conclusions that seem tenable today, it is reasonable to suppose that a generative grammar is a system of many hundreds of rules of several different types, organized in accordance with certain fixed principles of ordering and applicability and containing a certain fixed substructure which, along with the general principles of organization, is common to all languages. There is no a priori "naturalness" to such a system, any more than there is to the detailed structure of the visual cortex. No one who has given any serious thought to the problem of formalizing inductive procedures or "heuristic methods" is likely to set much store by the hope that such a system as a generative grammar can be constructed by methods of any generality.

To my knowledge, the only substantive proposal to deal with the problem of acquisition of knowledge of language is the rationalist conception that I have outlined. To repeat: suppose that we assign to the mind, as an innate property, the general theory of language that we have called "universal grammar." This theory encompasses the principles that I discussed in the preceding lecture and many others of the same sort, and it specifies a certain subsystem of rules that provides a skeletal structure for any language and a variety of conditions, formal and substantive, that any further elaboration of the grammar must meet. The theory of universal grammar, then, provides a schema to which any particular grammar must conform. Suppose, furthermore, that we can make this schema sufficiently restrictive so that very few possible grammars conforming to the schema will be consistent with the meager and degenerate data actually available to the language learner. His task, then, is to search among the possible grammars and select one that is not definitely rejected by the data available to him. What faces the language learner, under these assumptions, is not the impossible task of inventing a highly abstract and intricately structured theory on the basis of degenerate data, but rather the much more manageable task of determining whether these data belong to one or another of a fairly restricted set of potential languages.

[18] It is surprising to see that Putnam refers disparagingly to "vague talk of 'classes of hypotheses' – and 'weighting functions'" in the course of his discussion of "general learning strategies." For the moment, the latter is a mere phrase without any describable content. On the other hand, there is a substantial literature detailing the properties of the classes of hypotheses and weighting functions to which Putnam refers. Hence, the shoe seems to be on the other foot in this case.

The tasks of the psychologist, then, divide into several subtasks. The first is to discover the innate schema that characterizes the class of potential languages – that defines the "essence" of human language. This subtask falls to that branch of human psychology known as linguistics; it is the problem of traditional universal grammar, of contemporary linguistic theory. The second subtask is the detailed study of the actual character of the stimulation and the organism–environment interaction that sets the innate cognitive mechanism into operation. This is a study now being undertaken by a few psychologists, and it is particularly active right here in Berkeley. It has already led to interesting and suggestive conclusions. One might hope that such study will reveal a succession of maturational stages leading finally to a full generative grammar.[19]

A third task is that of determining just what it means for a hypothesis about the generative grammar of a language to be "consistent" with the data of sense. Notice that it is a great oversimplification to suppose that a child must discover a generative grammar that accounts for all the linguistic data that has been presented to him and that "projects" such data to an infinite range of potential sound–meaning relations. In addition to achieving this, he must also differentiate the data of sense into those utterances that give direct evidence as to the character of the underlying grammar and those that must be rejected by the hypothesis he selects as ill-formed, deviant, fragmentary, and so on. Clearly, everyone succeeds in carrying out this task of differentiation – we all know, within tolerable limits of consistency, which sentences are well formed and literally interpretable, and which must be interpreted as metaphorical, fragmentary, and deviant along many possible dimensions. I doubt that it has been fully appreciated to what extent this complicates the problem of accounting for language acquisition. Formally speaking, the learner must select a hypothesis regarding the language to which he is exposed that rejects a good part of the data on which this hypothesis must rest. Again, it is reasonable to suppose this is possible only if the range of tenable hypotheses is quite limited – if the innate schema of universal grammar is highly restrictive. The third subtask, then, is to study what we might think of as the problem of "confirmation" – in this context, the problem of what relation must hold between a potential grammar and a set

[19] It is not unlikely that detailed investigation of this sort will show that the conception of universal grammar as an innate schematism is only valid as a first approximation; that, in fact, an innate schematism of a more general sort permits the formulation of tentative "grammars" which themselves determine how later evidence is to be interpreted, leading to the postulation of richer grammars, and so on. I have so far been discussing language acquisition on the obviously false assumption that it is an instantaneous process. There are many interesting questions that arise when we consider how the process extends in time. For some discussion relating to problems of phonology, see my paper "Phonology and Reading," in H. Levin, ed., *Basic Studies on Reading.* Notice also that it is unnecessary to suppose, even in the first approximation, that "very few possible grammars conforming to the schema" will be available to the language learner. It is enough to suppose that the possible grammars consistent with the data will be "scattered" in terms of an evaluation procedure.

of data for this grammar to be confirmed as the actual theory of the language in question.

I have been describing the problem of acquisition of knowledge of language in terms that are more familiar in an epistemological than a psychological context, but I think that this is quite appropriate. Formally speaking, acquisition of "common-sense knowledge" – knowledge of a language, for example – is not unlike theory construction of the most abstract sort. Speculating about the future development of the subject, it seems to me not unlikely, for the reasons I have mentioned, that learning theory will progress by establishing the innately determined set of possible hypotheses, determining the conditions of interaction that lead the mind to put forth hypotheses from this set, and fixing the conditions under which such a hypothesis is confirmed – and, perhaps, under which much of the data is rejected as irrelevant for one reason or another.

Such a way of describing the situation should not be too surprising to those familiar with the history of psychology at Berkeley, where, after all, Edward Tolman has given his name to the psychology building; but I want to stress that the hypotheses I am discussing are qualitatively different in complexity and intricacy from anything that was considered in the classical discussions of learning. As I have now emphasized several times, there seems to be little useful analogy between the theory of grammar that a person has internalized and that provides the basis for his normal, creative use of language, and any other cognitive system that has so far been isolated and described; similarly, there is little useful analogy between the schema of universal grammar that we must, I believe, assign to the mind as an innate character, and any other known system of mental organization. It is quite possible that the lack of analogy testifies to our ignorance of other aspects of mental function, rather than to the absolute uniqueness of linguistic structure; but the fact is that we have, for the moment, no objective reason for supposing this to be true.

The way in which I have been describing acquisition of knowledge of language calls to mind a very interesting and rather neglected lecture given by Charles Sanders Peirce more than fifty years ago, in which he developed some rather similar notions about acquisition of knowledge in general.[20] Peirce argued that the general limits of human intelligence are much more narrow than might be suggested by romantic assumptions about the limitless perfectibility of man (or, for that matter, than are suggested by his own "pragmaticist" conceptions of the course of scientific progress in his better-known philosophical studies). He held that innate limitations on admissible hypotheses are a precondition for successful theory construction, and that the "guessing instinct" that provides hypotheses makes use of inductive procedures only for "corrective action."

[20] C. S. Peirce, "The Logic of Abduction," in V. Tomas, ed., *Peirce's Essays in the Philosophy of Science* (New York: Liberal Arts Press, 1957).

Peirce maintained in this lecture that the history of early science shows that something approximating a correct theory was discovered with remarkable ease and rapidity, on the basis of highly inadequate data, as soon as certain problems were faced; he noted "how few were the guesses that men of surpassing genius had to make before they rightly guessed the laws of nature." And, he asked, "How was it that man was ever led to entertain that true theory? You cannot say that it happened by chance, because the chances are too overwhelmingly against the single true theory in the twenty or thirty thousand years during which man has been a thinking animal, ever having come into any man's head." A fortiori, the chances are even more overwhelmingly against the true theory of each language ever having come into the head of every four-year-old child. Continuing with Peirce: "Man's mind has a natural adaptation to imagining correct theories of some kinds . . . If man had not the gift of a mind adapted to his requirements, he could not have acquired any knowledge." Correspondingly, in our present case, it seems that knowledge of a language – a grammar – can be acquired only by an organism that is "preset" with a severe restriction on the form of grammar. This innate restriction is a precondition, in the Kantian sense, for linguistic experience, and it appears to be the critical factor in determining the course and result of language learning. The child cannot know at birth which language he is to learn, but he must know that its grammar must be of a predetermined form that excludes many imaginable languages. Having selected a permissible hypothesis, he can use inductive evidence for corrective action, confirming or disconfirming his choice. Once the hypothesis is sufficiently well confirmed, the child knows the language defined by this hypothesis; consequently, his knowledge extends enormously beyond his experience and, in fact, leads him to characterize much of the data of experience as defective and deviant.

Peirce regarded inductive processes as rather marginal to the acquisition of knowledge; in his words, "Induction has no originality in it, but only tests a suggestion already made." To understand how knowledge is acquired, in the rationalist view that Peirce outlined, we must penetrate the mysteries of what he called "abduction," and we must discover that which "gives a rule to abduction and so puts a limit upon admissible hypotheses." Peirce maintained that the search for principles of abduction leads us to the study of innate ideas, which provide the instinctive structure of human intelligence. But Peirce was no dualist in the Cartesian sense; he argued (not very persuasively, in my opinion) that there is a significant analogy between human intelligence, with its abductive restrictions, and animal instinct. Thus, he maintained that man discovered certain true theories only because his "instincts must have involved from the beginning certain tendencies to think truly" about certain specific matters; similarly, "You cannot seriously think that every little chicken that is hatched, has to rummage through all possible theories until it lights upon the good idea of picking up something and eating it. On the contrary, you think that the chicken

has an innate idea of doing this; that is to say, that it can think of this, but has no faculty of thinking anything else . . . But if you are going to think every poor chicken endowed with an innate tendency towards a positive truth, why should you think to man alone this gift is denied?"

No one took up Peirce's challenge to develop a theory of abduction, to determine those principles that limit the admissible hypotheses or present them in a certain order. Even today, this remains a task for the future. It is a task that need not be undertaken if empiricist psychological doctrine can be substantiated; therefore, it is of great importance to subject this doctrine to rational analysis, as has been done, in part, in the study of language. I would like to repeat that it was the great merit of structural linguistics, as of Hullian learning theory in its early stages and of several other modern developments, to have given precise form to certain empiricist assumptions.[21] Where this step has been taken, the inadequacy of the postulated mechanisms has been clearly demonstrated, and, in the case of language at least, we can even begin to see just why any methods of this sort must fail – for example, because they cannot, in principle, provide for the properties of deep structures and the abstract operations of formal grammar. Speculating about the future, I think it is not unlikely that the dogmatic character of the general empiricist framework and its inadequacy to

[21] In contrast, the account of language acquisition presented by B. F. Skinner in his *Verbal Behavior* (New York: Appleton-Century-Crofts, 1957) seems to me either devoid of content or clearly wrong, depending on whether one interprets it metaphorically or literally (see my review of this book in *Language*, Vol. 35, No. 1, 1959, pp. 26–58). It is quite appropriate when a theory is disproven in a strong form to replace it by a weaker variant. However, not infrequently this step leads to vacuity. The popularity of Skinner's concept of "reinforcement," after the virtual collapse of Hullian theory, seems to me a case in point. (Note that the Skinnerian concepts can be well defined and can lead to interesting results, in a particular experimental situation – what is at issue is the Skinnerian "extrapolation" to a wider class of cases.)

Another example appears in K. Salzinger, "The Problem of Response Class in Verbal Behavior," in K. Salzinger and S. Salzinger, eds., *Research in Verbal Behavior and Some Neurophysiological Implications* (New York: Academic Press, 1967), pp. 35–54. Salzinger argues that George Miller is not justified in criticizing learning theory for its inability to explain linguistic productivity – that is, the ability of a speaker to determine, of a sequence of words that he has never heard, whether or not it is a well-formed sentence and what it means. The defect can be overcome, he argues, by making use of the notion of "response class." True, it cannot be that each response is reinforced, but the class of acceptable sentences constitutes a response class, like the set of bar-presses in a particular Skinnerian experiment. Unfortunately, this is empty verbiage until the condition that defines membership in this class is established. If the condition involves the notion "generation by a given grammar," then we are back where we started.

Salzinger also misconstrues the attempts to provide an experimental test that will distinguish grammatical from ungrammatical strings. He states that such tests have failed to confirm such a division and therefore concludes, apparently, that the distinction does not exist. Obviously, the failure indicates nothing more than that the tests were ineffective. One can invent innumerable tests that would fail to provide some given classification. Surely the classification itself is not in question. Thus, Salzinger would agree, quite apart from any experimental test that might be devised, that the sentences of this footnote share an important property that does not hold of the set of strings of words formed by reading each of these sentences, word by word, from right to left.

human and animal intelligence will gradually become more evident as specific realizations, such as taxonomic linguistics, behaviorist learning theory, and the perception models,[22] heuristic methods, and "general problem solvers" of the early enthusiasts of "artificial intelligence," are successively rejected on empirical grounds when they are made precise and on grounds of vacuity when they are left vague. And – assuming this projection to be accurate – it will then be possible to undertake a general study of the limits and capacities of human intelligence, to develop a Peircean logic of abduction.

Modern psychology is not devoid of such initiatives. The contemporary study of generative grammar and its universal substructure and governing principles is one such manifestation. Closely related is the study of the biological bases of human language, an investigation to which Eric Lenneberg has made substantial contributions.[23] It is tempting to see a parallel development in the very important work of Piaget and others interested in "genetic epistemology," but I am not sure that this is accurate. It is not clear to me, for example, what Piaget takes to be the basis for the transition from one of the stages that he discusses to the next, higher stage. There is, furthermore, a possibility, suggested by recent work of Mehler and Bever,[24] that the deservedly well-known results on conservation, in particular, may not demonstrate successive stages of intellectual development in the sense discussed by Piaget and his coworkers, but something rather different. If the preliminary results of Mehler and Bever are correct, then it would follow that the "final stage," in which conservation is properly understood, was already realized at a very early period of development. Later, the child develops a heuristic technique that is largely adequate but that fails under the conditions of the conservation experiment. Still later, he adjusts this technique successfully and once again makes the correct judgments in the conservation experiment. If this analysis is correct, then what we are observing is not a succession of stages of intellectual development, in Piaget's sense, but rather slow progress in bringing heuristic techniques into line with general concepts that have always been present. These are interesting alternatives; either way, the results may bear in important ways on the topics we are considering.

Still more clearly to the point, I think, are the developments in comparative ethology over the past thirty years, and certain current work in experimental and physiological psychology. One can cite many examples: for example, in the

[22] For a discussion of such systems and their limitations, see M. Minsky and S. Papert, *Perceptions and Pattern Recognition*, Artificial Intelligence Memo No. 140, MAC-M-358, Project MAC, Cambridge, Mass., September 1967.

[23] See E. H. Lenneberg, *Biological Foundations of Language* (New York: Wiley, 1967). My contribution to this volume, "The Formal Nature of Language," appears as the fifth paper in this book.

[24] See J. Mehler and T. G. Bever, "Cognitive Capacities of Young Children," *Science*, Vol. 158, No. 3797, October 1967, pp. 141–42.

latter category, the work of Bower suggesting an innate basis for the perceptual constancies; studies in the Wisconsin primate laboratory on complex innate releasing mechanisms in rhesus monkeys; the work of Hubel, Barlow, and others on highly specific analyzing mechanisms in the lower cortical centers of mammals; and a number of comparable studies of lower organisms (for example, the beautiful work of Lettvin and his associates on frog vision). There is now good evidence from such investigations that perception of line, angle, motion, and other complex properties of the physical world is based on innate organization of the neural system.

In some cases at least, these built-in structures will degenerate unless appropriate stimulation takes place at an early stage in life, but although such experience is necessary to permit the innate mechanisms to function, there is no reason to believe that it has more than a marginal effect on determining *how* they function to organize experience. Furthermore, there is nothing to suggest that what has so far been discovered is anywhere near the limit of complexity of innate structures. The basic techniques for exploring the neural mechanisms are only a few years old, and it is impossible to predict what order of specificity and complexity will be demonstrated when they come to be extensively applied. For the present, it seems that most complex organisms have highly specific forms of sensory and perceptual organization that are associated with the *Umwelt* and the manner of life of the organism. There is little reason to doubt that what is true of lower organisms is true of humans as well. Particularly in the case of language, it is natural to expect a close relation between innate properties of the mind and features of linguistic structure; for language, after all, has no existence apart from its mental representation. Whatever properties it has must be those that are given to it by the innate mental processes of the organism that has invented it and that invents it anew with each succeeding generation, along with whatever properties are associated with the conditions of its use. Once again, it seems that language should be, for this reason, a most illuminating probe with which to explore the organization of mental processes.

Turning to comparative ethology, it is interesting to note that one of its earliest motivations was the hope that through the "investigation of the a priori, of the innate working hypotheses present in subhuman organisms," it would be possible to shed light on the a priori forms of human thought. This formulation of intent is quoted from an early and little-known paper by Konrad Lorenz.[25] Lorenz goes on to express views very much like those Peirce had expressed a generation earlier. He maintains:

[25] K. Lorenz, "Kants Lehre vom apriorischen in Lichte gegenwärtiger Biologie," in *Blätter für Deutsche Philosophie*, Vol. 15, 1941, pp. 94–125. I am indebted to Donald Walker of the MITRE Corporation, Bedford, Mass., for bringing this paper to my attention.

> One familiar with the innate modes of reaction of subhuman organisms can readily hypothesize that the a priori is due to hereditary differentiations of the central nervous system which have become characteristic of the species, producing hereditary dispositions to think in certain forms . . . Most certainly Hume was wrong when he wanted to derive all that is a priori from that which the senses supply to experience, just as wrong as Wundt or Helmholtz who simply explain it as an abstraction from preceding experience. Adaptation of the a priori to the real world has no more originated from "experience" than adaptation of the fin of the fish to the properties of water. Just as the form of the fin is given a priori, prior to any individual negotiation of the young fish with the water, and just as it is this form that makes possible this negotiation, so it is also the case with our forms of perception and categories in their relationship to our negotiation with the real external world through experience. In the case of animals, we find limitations specific to the forms of experience possible for them. We believe we can demonstrate the closest functional and probably genetic relationship between these animal a prioris and our human a priori. Contrary to Hume, we believe, just as did Kant, that a "pure" science of innate forms of human thought, independent of all experience, is possible.

Peirce, to my knowledge, is original and unique in stressing the problem of studying the rules that limit the class of possible theories. Of course, his concept of abduction, like Lorenz's biological a priori, has a strongly Kantian flavor, and all derive from the rationalist psychology that concerned itself with the forms, the limits, and the principles that provide "the sinews and connections" for human thought, that underlie "that infinite amount of knowledge of which we are not always conscious," of which Leibnitz spoke. It is therefore quite natural that we should link these developments to the revival of philosophical grammar, which grew from the same soil as an attempt, quite fruitful and legitimate, to explore one basic facet of human intelligence.

In recent discussion, models and observations derived from ethology have frequently been cited as providing biological support, or at least analogue, to new approaches to the study of human intelligence. I cite these comments of Lorenz's mainly in order to show that this reference does not distort the outlook of at least some of the founders of this domain of comparative psychology.

One word of caution is necessary in referring to Lorenz, now that he has been discovered by Robert Ardrey and Joseph Alsop and popularized as a prophet of doom. It seems to me that Lorenz's views on human aggression have been extended to near absurdity by some of his expositors. It is no doubt true that there are innate tendencies in the human psychic constitution that lead to aggressiveness under specific social and cultural conditions. But there is little reason to suppose that these tendencies are so dominant as to leave us forever tottering on the brink of a Hobbesian war of all against all – as, incidentally, Lorenz at least is fully aware, if I read him rightly. Skepticism is certainly in order when a doctrine of man's "inherent aggressiveness" comes to the surface in a society that glorifies competitiveness, in a civilization that has

been distinguished by the brutality of the attack that it has mounted against less fortunate peoples. It is fair to ask to what extent the enthusiasm for this curious view of man's nature is attributable to fact and logic and to what extent it merely reflects the limited extent to which the general cultural level has advanced since the days when Clive and the Portuguese explorers taught the meaning of true savagery to the inferior races that stood in their way.

In any event, I would not want what I am saying to be confused with other, entirely different attempts to revive a theory of human instinct. What seems to me important in ethology is its attempt to explore the innate properties that determine how knowledge is acquired and the character of this knowledge. Returning to this theme, we must consider a further question: how did the human mind come to acquire the innate structure that we are led to attribute to it? Not too surprisingly, Lorenz takes the position that this is simply a matter of natural selection. Peirce offers a rather different speculation, arguing that "nature fecundates the mind of man with ideas which, when these ideas grow up, will resemble their father, Nature." Man is "provided with certain natural beliefs that are true" because "certain uniformities . . . prevail throughout the universe, and the reasoning mind is [it]self a product of this universe. These same laws are thus, by logical necessity, incorporated in his own being." Here, it seems clear that Peirce's argument is entirely without force and that it offers little improvement over the preestablished harmony that it was presumably intended to replace. The fact that the mind is a product of natural laws does not imply that it is equipped to understand these laws or to arrive at them by "abduction." There would be no difficulty in designing a device (say, programing a computer) that is a product of natural law, but that, given data, will arrive at any arbitrary absurd theory to "explain" these data.

In fact, the processes by which the human mind achieved its present stage of complexity and its particular form of innate organization are a total mystery, as much so as the analogous questions about the physical or mental organization of any other complex organism. It is perfectly safe to attribute this development to "natural selection," so long as we realize that there is no substance to this assertion, that it amounts to nothing more than a belief that there is some naturalistic explanation for these phenomena. The problem of accounting for evolutionary development is, in some ways, rather like that of explaining successful abduction. The laws that determine possible successful mutation and the nature of complex organisms are as unknown as the laws that determine the choice of hypotheses.[26] With no knowledge of the laws that determine the

[26] It has been argued on statistical grounds – through comparison of the known rate of mutation with the astronomical number of imaginable modifications of chromosomes and their parts – that such laws must exist and must vastly restrict the realizable possibilities. See the papers by Eden, Schützenberger, and Gavadan in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, Wistar Symposium Monograph No. 5, 1967.

organization and structure of complex biological systems, it is just as senseless to ask what the "probability" is for the human mind to have reached its present state as it is to inquire into the "probability" that a particular physical theory will be devised. And, as we have noted, it is idle to speculate about laws of learning until we have some indication of what kind of knowledge is attainable – in the case of language, some indication of the constraints on the set of potential grammars.

In studying the evolution of mind, we cannot guess to what extent there are physically possible alternatives to, say, transformational generative grammar, for an organism meeting certain other physical conditions characteristic of humans. Conceivably, there are none – or very few – in which case talk about evolution of the language capacity is beside the point. The vacuity of such speculation, however, has no bearing one way or another on those aspects of the problem of mind that can be sensibly pursued. It seems to me that these aspects are, for the moment, the problems illustrated in the case of language by the study of the nature, the use, and the acquisition of linguistic competence.

There is one final issue that deserves a word of comment. I have been using mentalistic terminology quite freely, but entirely without prejudice as to the question of what may be the physical realization of the abstract mechanisms postulated to account for the phenomena of behavior or the acquisition of knowledge. We are not constrained, as was Descartes, to postulate a second substance when we deal with phenomena that are not expressible in terms of matter in motion, in his sense. Nor is there much point in pursuing the question of psychophysical parallelism, in this connection. It is an interesting question whether the functioning and evolution of human mentality can be accommodated within the framework of physical explanation, as presently conceived, or whether there are new principles, now unknown, that must be invoked, perhaps principles that emerge only at higher levels of organization than can now be submitted to physical investigation. We can, however, be fairly sure that there will be a physical explanation for the phenomena in question, if they can be explained at all, for an uninteresting terminological reason, namely that the concept of "physical explanation" will no doubt be extended to incorporate whatever is discovered in this domain, exactly as it was extended to accommodate gravitational and electromagnetic force, massless particles, and numerous other entities and processes that would have offended the common sense of earlier generations. But it seems clear that this issue need not delay the study of the topics that are now open to investigation, and it seems futile to speculate about matters so remote from present understanding.

I have tried to suggest that the study of language may very well, as was traditionally supposed, provide a remarkably favorable perspective for the study of human mental processes. The creative aspect of language use, when investigated with care and respect for the facts, shows that current notions of habit

and generalization, as determinants of behavior or knowledge, are quite inadequate. The abstractness of linguistic structure reinforces this conclusion, and it suggests further that in both perception and learning the mind plays an active role in determining the character of the acquired knowledge. The empirical study of linguistic universals has led to the formulation of highly restrictive and, I believe, quite plausible hypotheses concerning the possible variety of human languages, hypotheses that contribute to the attempt to develop a theory of acquisition of knowledge that gives due place to intrinsic mental activity. It seems to me, then, that the study of language should occupy a central place in general psychology.

Surely the classical questions of language and mind receive no final solution, or even the hint of a final solution, from the work that is being actively pursued today. Nevertheless, these problems can be formulated in new ways and seen in a new light. For the first time in many years, it seems to me, there is some real opportunity for substantial progress in the study of the contribution of the mind to perception and the innate basis for acquisition of knowledge. Still, in many respects, we have not made the first approach to a real answer to the classical problems. For example, the central problems relating to the creative aspect of language use remain as inaccessible as they have always been. And the study of universal semantics, surely crucial to the full investigation of language structure, has barely advanced since the medieval period. Many other critical areas might be mentioned where progress has been slow or nonexistent. Real progress has been made in the study of the mechanisms of language, the formal principles that make possible the creative aspect of language use and that determine the phonetic form and semantic content of utterances. Our understanding of these mechanisms, though only fragmentary, does seem to me to have real implications for the study of human psychology. By pursuing the kinds of research that now seem feasible and by focusing attention on certain problems that are now accessible to study, we may be able to spell out in some detail the elaborate and abstract computations that determine, in part, the nature of percepts and the character of the knowledge that we can acquire – the highly specific ways of interpreting phenomena that are, in large measure, beyond our consciousness and control and that may be unique to man.