PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Knowledge of Language: Its Elements and Origins

N. Chomsky

Phil. Trans. R. Soc. Lond. B 1981 **295**, 223-234
doi: 10.1098/rstb.1981.0135

Phil. Trans. R. Soc. Lond. B **295**, 223–234 (1981)
Printed in Great Britain

Knowledge of language: its elements and origins

By N. Chomsky, Corresponding F.B.A.

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, *U.S.A.*

My approach to the study of language is based on the assumption that knowledge of language can be properly characterized by means of a generative grammar, i.e. a system of rules and principles that assigns structural descriptions to linguistic expressions. On this view, the basic concepts are those of 'grammar' and 'knowledge of grammar'. The concepts of 'language' and 'knowledge of language' are derivative: they involve a higher level of abstraction from psychological mechanisms and raise additional (though not necessarily important) problems.

Of central concern, from this point of view, will be to determine the biological endowment that makes it possible for a grammar of the required sort to develop in human beings provided that they are exposed to some appropriate body of experience. This biological endowment may be regarded as a function that maps a body of experience into a particular grammar. The function itself is commonly referred to as universal grammar (u.g.) and can be expressed, in part, as a system of principles that determine the class of accessible particular grammars and their properties. Recent work suggests that u.g. consists, on the one hand, of a theory of so-called core grammar and, on the other, of a theory of permissible extensions and modifications of core grammar. Given the intricate internal structure of u.g., it can account for the superficially highly diverse grammars and languages that do in fact exist. Thus, what appear to be quite different systems of knowledge may arise from relatively little experience.

A number of subsystems of u.g. have now been explored, each with its distinctive properties and possibilities of variation. Some current proposals concerning these systems are sketched, and some consequences considered with regard to the nature and acquisition of cognitive systems (including systems of knowledge) more generally.

Grammar and universal grammar

Three questions arise at once when we consider the notion 'knowledge of language': (I) What is the nature of this knowledge? (II) How is it acquired? (III) How is it put to use? The first question has a certain conceptual priority; we can hope to gain some insight into acquisition or use of knowledge only to the extent that we have some understanding of what it is that is acquired or used.

A standard, and I think basically correct, approach to the first question is to assume that a person who knows a language has internalized a grammar, a system of rules and principles that assigns structural descriptions to linguistic expressions. If so, then the central notions to be developed are the correlative notions 'grammar' and 'structural description', and the central cognitive relation is the one that holds between the person and the grammar. I shall refer to this relation as '(tacit) knowledge', thus saying that a person knows his grammar and knows the rules and principles of his grammar, which of course does not imply that he has propositional knowledge that these are the rules and principles of his grammar. The linguist may develop such propositional knowledge, but that is quite a different matter.

A person who knows English has attained a certain mental state, different from that of someone who knows Japanese. Abstracting from possible individual differences, there is some innate mental state common to the species that provides the basis for acquisition of knowledge of grammar, a characteristic that distinguishes humans from birds or apes. One proposal, which I think is basically correct, is that this innate endowment consists of a system of principles, each with certain possibilities of parametric variation, and that acquisition of knowledge of grammar with all that it entails is, in part, a matter of setting these parameters one way or another on the basis of presented experience. Let us say that this process yields a 'core grammar'. Then an actual grammar, representing full knowledge of language, consists of a core grammar extended to a periphery that incorporates more idiosyncratic (marked) elements, also constrained by innate principles but in somewhat different ways. There are familiar idealizations implicit in this rough account, but they are, I think, quite legitimate ones.

One component of the human mind–brain, then, is a genetically determined initial configuration, which we may call 'the initial state of the language faculty'. It is characterized by a theory of principles and parameters and a theory of markedness, which permits the extension of core grammar to a full grammar. I shall refer to this complex as 'a theory of universal grammar (u.g.)', and shall say that u.g. is a component of the initial state. Under the triggering and shaping effect of experience, the initial state is transformed to a more or less steady state incorporating the attained grammar. My use of mentalistic terminology involves nothing beyond the assumption that u.g. and grammar are elements of the initial and steady state, respectively, characterized at some appropriate level of abstraction; in short, that this characterization expresses conditions satisfied in some way by physical mechanisms of the brain. In this usage, we would, for example, refer to the proposals of David Marr and his colleagues concerning visual mechanisms as 'mentalistic', for example the rigidity principle that permits determination of structure from motion, or the theory of visual representation in terms of axes of elongation, etc. (Marr & Nishihara 1978; Ullman 1979).

Another familiar idealization is the assumption that the transition from initial to steady state can be regarded as instantaneous, i.e. that the actual interstate transitions have at most a negligible effect on the state attained. This is surely questionable, but so far as is known it yields an extremely good first approximation, and I shall adopt it here. We may then think of u.g. as a function mapping presented experience to steady state attained, through the means of fixing parameters of u.g. and adding a marked periphery.

U.G. AS AN EXPLANATORY THEORY

Approaching our initial questions in this way, the second question is in part answered by the theory of u.g. Furthermore, this account of growth of knowledge can be interpreted as a model of explanation. There is a vast array of facts about knowledge of language that we would like to explain. Consider, for example, the sentences (1), (2).

(1) everyone wants John to like him
(2) everyone wants him to be liked by John

The sentence (1) can have the interpretation expressed in (3), whereas (2) cannot have the corresponding interpretation expressed in (4).

(3) for every person x, x wants John to like x
(4) for every person x, x wants x to be liked by John

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Thus, even though 'John likes him' and 'he is liked by John' are virtual synonyms, insertion of the corresponding infinitives as the clausal complement of the verb *want* yields expressions that differ in their range of meaning. The language-user's knowledge of these facts is propositional knowledge that so-and-so.

Such examples of propositional knowledge extend beyond the bounds of well-formed English sentences. It is, for example, a property of my dialect of English that the verb *want* does not take a *that*-clause complement, e.g. the sentence (5) is not well-formed.

(5) everyone wants that John like him

Nevertheless, I know that this expression, were it well-formed, could have the meaning expressed in (6), just as (7) – in contrast to (2) – could have the meaning expressed in (8).

(6) for every person x, x wants that John like x
(7) everyone wants that he be liked by John
(8) for every person x, x wants that x be liked by John

In this case, interchange of synonyms ('John like him', 'he be liked by John') does preserve meaning.

How can we explain these facts? At one level, we can explain them by postulating a grammar that yields these consequences, taking it to be a component of the steady state attained. At a deeper level, we can explain these facts by postulating a u.g. as an element of the initial state with the following property: given experience sufficient for language acquisition, this u.g. determines a grammar that in turn yields these consequences. A theory of u.g. with this property is a genuine explanatory theory. In the case illustrated, explanatory adequacy in this sense can be partly achieved. I shall not formulate the relevant principles of u.g. here; let us refer to them as 'the binding theory'. There has been a good deal of work in the past 10 years on binding theory, and principles have been formulated that are quite simple and natural and that have explanatory force over a fairly broad empirical domain (see Chomsky 1981*a*, *b*).

The binding theory entails that certain configurations are 'opaque' in the sense that an anaphor such as *each other* must be bound within them. An element is bound if it has an antecedent that c-commands it, where in the simplest case a category A c-commands B if A is an immediate constituent of a category that contains B but A does not contain B; then B is in the c-command domain of A. The opaque configurations include the subject of a finite clause and the minimal c-command domain of the subject of any clause. A pronominal such as *he* must be free exactly where an anaphor must be bound. By using indices in the obvious way to express the binding relation, typical examples of the opacity conditions are the following.

(9) John_i expects [Bill_j to like him_k]
(10) John_i expects [him_j to win]
(11) John_i expects [that he_j will win]

The brackets bound embedded clauses, finite in (11) and infinitival in (9), (10). In (9), *him* is in the minimal c-command domain of the subject *Bill*, and since it must be free in this domain, $k \neq j$. But k may $= i$. In (10), $j \neq i$ or *him* will be bound in the minimal c-command domain of the subject *John*. In (11), j may $= i$ since in any event *he* is free as subject of a finite clause. Anaphors satisfy the same conditions, with 'free' replaced by 'bound'. Coindexing may be interpreted here as intended coreference. Thus in (9) we may understand *him* as referring to

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

John but not to Bill. It is easy to show that the examples discussed earlier fall under the opacity conditions, where we interpret a pronoun coindexed with a variable as an occurrence of this variable (see Higginbotham 1980).

Another consequence of the binding theory is that names cannot be bound by pronouns. Thus we can have the binding represented in (12) but not (13).

(12) the woman he$_i$ married expected [us to like John$_i$]

(13) he$_i$ expected [us to like John$_i$]

In (12), *he* does not c-command *John*, so *John* is free, though *he* is its antecedent. But in (13) *he* does c-command *John* so that the coindexing violates the binding theory.

We can account for the propositional knowledge that these sentences have the interpretations indicated on the assumption that the binding theory is an innate property (possibly parametrized) and that experience provides sufficient evidence to determine that *he* is a pronoun. Many other cases are similar.

Consideration of the binding theory provides extensive evidence concerning the form of mental representations of linguistic expressions. Consider, for example, sentence (14).

(14) who did John$_i$ expect to like him$_k$

Here *him* may be understood to refer to John, so that k may $= i$. This fact is compatible with the binding theory only if *him* is in the minimal c-command domain of the subject of an embedded clausal complement of *expect*. No such subject is overtly expressed, but we can account for our knowledge of the facts on the assumption that such a subject is mentally represented; that is, at the level of syntactic representation at which the binding theory applies, the representation of (14) must be (15).

(15) who$_j$ did John$_i$ expect [x_j to like him$_k$]

Here x is the variable bound by the quasi-quantifier *who*. Since this mentally represented variable is the subject of the embedded clause, the binding relation between *John* and *him* satisfies the binding theory if $k = i$. The representation (15) is quite natural; with $k = i$, we may assume it to correspond to the logical form (16).

(16) for which person x, John$_i$ expected [x to like him$_i$]

The clause now has the form of (9), with $k = i$. Note that we cannot set $k = j$ in (15), thus replacing *him* in (16) by x, or the binding theory will be violated, just as *him* cannot refer to John in 'John likes him', 'I expect John to like him', etc.

The example (14) illustrates the fact that a subject may be a mentally represented empty element, lacking phonetic content. There is, incidentally, evidence that such empty elements, though lacking phonetic content, may affect the phonetic form of utterances (see Chomsky 1980, 1981*b*). Thus we have independent phonetic evidence that such empty elements are mentally represented; the phonetic rules actually 'see' them.

Other examples show that a direct object may also be a mentally represented empty element. Consider the sentence (17).

(17) I wonder whom he expected us to like

We cannot understand (17) as having the meaning expressed in (18).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(18) I wonder for which person x, x expected us to like x

In contrast, (18) can be the logical form corresponding to (19).

(19) I wonder who expected us to like him

These facts follow if we assume the syntactic representations of (17) and (19) to be (20) and (21), respectively.

(20) I wonder [for which person x, he expected [us to like x]]

(21) I wonder [for which person x, x expected [us to like him]]

The embedded complement of *wonder* in (20) has essentially the form of (13), so that the variable x cannot be bound by *he* just as *John* cannot be bound by *he* in (13), accounting for the fact that (18) is not the interpretation of (17). No binding principle prevents (18) from being the interpretation of (19) or (21), however. In this case, the embedded complement of *wonder* has the form of (9). Note that these examples provide evidence that variables behave in the manner of names, with regard to the binding theory.

These examples illustrate the fact that either a subject or an object may be a mentally represented empty element. More complex cases show that both may be empty elements, as we would expect. Consider the sentences (22), (23).

(22) John is too stubborn to talk to Bill

(23) John is too stubborn to talk to

We understand these sentences as (24), (25), respectively.

(24) John_i is so stubborn that he_i will not talk to Bill

(25) John_i is so stubborn that one cannot talk to him_i

What is particularly interesting about these examples is that the understood subject of the transitive verb *talk to* is interpreted differently in the two cases: it is taken to refer to John in (22) but to some arbitrary person in (23). Yet these sentences differ only in that the object of *talk to* is overtly expressed in (22), while it is missing in (23). These curious facts also follow from the binding theory, if we assume that the 'understood subject' and 'understood object' are in fact mentally represented, as in (26) and (27), corresponding respectively to (22) and (23).

(26) John is too stubborn [PRO to talk to Bill]

(27) John_i is too stubborn [PRO_j to talk to X_k]

What I have represented as 'PRO' is to be understood as an abstract pronominal, i.e. a pronoun lacking phonetic content. The binding theory permits PRO to be bound by *John* in both (26) and (27), and another subtheory, the theory of control, requires this binding in (26). Turning to (27), the binding theory prevents binding of X by PRO: thus $k \neq j$, exactly as in (9). Another consequence of the binding theory, which there is no space to explain here, is that X in (27) cannot be free PRO; since $k \neq j$, it must be that $k = i$ since X cannot be free. Therefore $j \neq i$. Since there is no other antecedent for PRO, it must be interpreted as referring to some arbitrary person, just as in (28), mentally represented as in (29).

(28) it is unclear how to solve the problem (to help oneself)

(29) it is unclear [how PRO to solve the problem (to help oneself)]

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Note that the reflexive option again entails that an antecedent must be present, namely, the empty pronominal PRO.

The careful reader should now be asking why PRO in (26) is permitted to be bound by *John*, even though it is a pronominal in the minimal c-command domain of this subject. This fact too follows from the binding theory, on the quite natural assumption that PRO, as distinct from an overt pronoun, is a pronominal anaphor. For details, and discussion of assumptions that are tacit here, see Chomsky (1981 *b*).

Further examples provide additional support for this analysis. Consider (30) and (31).

(30) John is too clever to expect us to catch Bill
(31) John is too clever to expect us to catch

In (30), we understand the subject of *expect* to be *John*; in (31), to be some arbitrary person. These facts again follow if we take the mental representations to be (32) and (33), respectively.

(32) John is too clever [PRO to expect [us to catch Bill]]
(33) John is too clever [PRO to expect [us to catch *X*]]

In (32), PRO can be coindexed with *John*. Turning to (33), as noted earlier, the binding theory requires that *X* be distinct from free PRO. Example (33), in fact, provides direct evidence that $X \neq$ PRO. For if *X* were PRO, then there would be no reason why it should not be coindexed with the PRO subject of *expect*, yielding an instance of the same configuration as (9), with $k = i$. But this interpretation is impossible, so *X* cannot be PRO. The only alternative is that *X* is a variable; for independent evidence in support of this conclusion, see Chomsky (1977, 1980, 1981 *a*, *b*). But this variable lacks an appropriate binding operator, so it must be assigned an antecedent. By the binding theory, its antecedent cannot be *us* or the PRO subject of *expect*. Therefore it must be *John*. Exactly as in case (27), it follows that PRO cannot be coindexed with *John*, and must therefore be arbitrary in interpretation.

The very simple principles of the binding theory that account for the behaviour of overt elements as in (9) and (11) also explain the properties of these considerably more complex examples, on the assumption that empty elements appear in mental representation where they are 'understood'. This fact provides evidence that such empty elements do appear in mental representations. As we have seen, these empty elements fall into two distinct types: variables, which behave in the manner of names, and PRO, which is a pronominal anaphor. There is in fact a third type, so-called 'trace', which is a pure anaphor (see Chomsky 1981 *b*). There is extensive further evidence that empty elements appear in syntactic representations exactly where they are 'understood' in the intuitive sense, and that these elements have quite specific properties and fall under three distinct types: anaphor, pronominal anaphor, and name-like variable. The evidence is similar to what I have just sketched: on these assumptions, a wide variety of empirical facts can be explained by the binding theory, which is independently motivated for overt elements, and other subtheories of u.g. Furthermore, the positions in which these various types of empty elements may appear are determined by the interaction of these subtheories. Finally, slight modifications in the parameters of the subtheories yields a substantially different array of data, which is attested in other languages. For example, a slight change in the parameter associated with one morphological rule yields the quite different properties of the so-called 'pro-drop languages' such as Italian and Spanish (see Chomsky (1981 *b*) and, for a slightly different approach, Rizzi (1980)). These are the kinds of results that we hope to attain in the study of u.g.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The projection principle

Examples of the sort just sketched suggest a very strong and quite natural principle, which I shall call 'the projection principle', as a component of u.g. The principle states that syntactic structure is projected from lexical properties in the sense that the argument structure of lexical items is represented explicitly at each syntactic level. Thus, the verbs *hit*, *help* and *talk to*, as a lexical property, take an object to which they assign a certain semantic role and a subject to which they assign a different semantic role. By the projection principle, at every syntactic level there must appear a subject and an object in the appropriate structural configuration. These appear overtly in the sentence 'John hits Bill' and they appear as empty elements in (26) and (27). I cannot give a precise formulation here (see Chomsky 1981*b*), but this is the intuitive idea.

The projection principle has a wide range of consequences. Assuming it to be a component of u.g., i.e. of innate endowment, a child who is 'learning English' somehow discovers the lexical properties of the verbs *hit*, *help* and *talk to* and then knows without further evidence that the examples just discussed are to be understood as indicated. The projection principle thus vastly facilitates the task of what is misleadingly called 'language learning' – a better term would be 'growth of grammar'. Equivalently, it has substantial explanatory force, as in the examples indicated. One virtual consequence of the projection principle is what has been called 'trace theory', which itself has substantial explanatory scope.

Turning briefly to question (III), – how is knowledge put to use? – the projection principle also has suggestive consequences. Equipped with this principle, and knowing lexical entries, a person who hears examples of the sort illustrated can at once construct the abstract representations that underlie them, yielding the required interpretations, given the binding theory and other subtheories of u.g.

Internal to the theory of grammar, the projection principle has the consequence that the complex apparatus of phrase structure rules can be eliminated, apart from language-specific parameters such as order of major categories. At the same time, it seems that the transformational component can be restricted to a single rule of core grammar – namely, the rule 'move any category anywhere' – with a few simple parameters. The grammar can therefore be quite 'small', containing few and simple rules, a property that itself has significant consequences for parsing and acquisition.

The single transformational rule of core grammar, which has broad scope, can be characterized abstractly in terms of several distinctive properties: it relates an empty element to an antecedent in a position assigned no semantic role; the empty element must be 'governed' in a sense that generalizes the traditional notion of government; and a strong locality principle holds of the antecedent–empty element relation. The empty pronominal PRO, in contrast, has quite different properties, which follow from the binding theory.

Let us explore some further consequences of the projection principle. Consider the sentence (34).

(34) the men asked me [how they could help each other]

Clearly, the bracketed phrase is clausal; it has the semantic properties of a clause and is the domain of *wh*-movement, which applies to clauses only. The verb *ask* takes an object and a clausal complement, so the projection principle is satisfied by the bracketing indicated.

Consider now the sentence (35).

(35) the men asked me [how to help each other]

Clearly, (35) is the counterpart to (34), with an infinitival rather than a finite clausal complement to *ask*, these being the two possible forms of a clause. Again, the semantic content is propositional and the complement is the domain of *wh*-movement. By the projection principle, the structure must be as in (36).

(36) the men asked me [how PRO to help each other]

The binding theory is satisfied with PRO as antecedent of *each other*. The surface form of (36) is derived by trivial rules, and an appropriate representation for logical form can also be derived in a straightforward way (see Higginbotham 1981).

Examining the distribution of the empty pronominal PRO, we discover some interesting properties. First, its distribution is virtually complementary to that of overt anaphors; there is (virtually) no context in which both can appear. This property is illustrated by interchange of PRO and *each other* in (36), which yields the non-sentence (37).

(37) the men asked me how each other to help PRO

This expression does not mean that each of the men asked me how the other is to help him, as the interpretive principles for the reciprocal imply, though this is a perfectly sensible interpretation of some sentence.

Secondly, we find that at a somewhat more abstract level of structure, the empty pronominal shares a good part of the distribution of anaphors, which is not surprising, given that it functions as a pronominal anaphor as already noted. For example, both can occur as subjects of infinitival clauses or of noun phrases, as in (36), (38) and (39), but not as subjects of tensed clauses as in (40).

(38) the men prefer [PRO helping each other]
(39) (i) the men would prefer [for each other to win]
(ii) the men would prefer [each other's books]
(40) (i) the men preferred [that PRO win the race]
(ii) the men preferred [that each other win the race]

Note that although PRO and *each other* share the distribution illustrated at an abstract level of structure, these examples still illustrate the complementary distribution in actual sentences; thus PRO and *each other* cannot be interchanged in the well-formed examples. There is, in fact, a restricted class of contexts in which both PRO and *each other* can appear, but it is reasonable to assume that this fact reflects an idiosyncratic marked property of English, and that at a more abstract level corresponding to core grammar the complementary distribution is not virtual but complete. As usual, the unanalysed phenomena that we observe tell us little in themselves.

Assuming this, we have to explain why PRO and *each other* are in complementary distribution in the class of actual sentential contexts, while they share fundamental distributional properties at a more abstract level. A theory of u.g. that does not yield this conclusion is plainly missing something important. The natural conclusion is that these facts reflect the interaction of two subsystems of u.g., one of which determines the general distribution of anaphors (both *each other* and PRO), and the second of which distinguishes overt from empty elements. In fact, the binding

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

theory yields the first of these results, as is fairly evident from the examples already discussed; and the theory of abstract case (in the sense of J.-R. Vergnaud) yields the second (see Chomsky 1981*a*, *b*).

This is a typical example of the modularity of structure of the language faculty, an internal counterpart to the modularity of the total system of cognitive structure. The latter assumption is, in my opinion, more controversial than it should be; it appears to be increasingly well supported as we learn more about particular cognitive systems, as well as quite natural. There is little reason to expect to find anything analogous to the rigidity principle in the theory of language, or anything analogous to the binding theory or the projection principle in the theory of vision. And despite much talk of generalized learning mechanisms, general principles of organism–environment interaction, and the like, I am unaware of any proposal substantive enough to bear investigation. On the basis of what is now known or plausibly surmised, there seems to be good reason to adopt the position that John Marshall has called 'the new organology', and to search, as he has suggested, for the neural correlates to this structural organization of mind (see Marshall 1980).

As the examples cited suggest, empty elements have provided an important probe into the properties of u.g. They are interesting from another point of view as well. Evidently, the grammar that is attained in the steady state results from the interaction of experience and initial state. It is difficult to believe that the intricate properties of empty elements are learned on the basis of direct evidence; in fact, one has no evidence bearing directly on their nature and properties. Therefore, it is reasonable to suppose that these properties reflect intrinsic properties of the mind–brain. If our interest is in the nature of the human mind–brain, rather than in phenomena of the external environment, then these elements have a particular fascination.

The growth of knowledge

Let us assume now that a successful theory of u.g. can be constructed along the lines indicated and think about the possible consequences. Note first that in the examples cited, and innumerable others like them, there is strong reason to believe that what we know, we know without relevant evidence. It is quite unreasonable to suppose that everyone who knows these facts has been given a sufficient basis of evidence to ground this knowledge. In many cases, it is difficult even to imagine what evidence might suffice for this purpose. There is no reason to believe that such propositional knowledge – for example, our knowledge that (31) means (33) – can be attained from evidence available to each speaker of the language by some generally valid procedure of induction or confirmation. Attention to the facts quickly convinces us that such an assumption is about as plausible as the belief that there exists some comparable general account for the fact that the human embryo grows a mammalian rather than an insect eye, or that children undergo puberty at a certain age, or that they employ the rigidity principle to determine structure from motion. In all such cases, it seems highly likely that innate properties account for the resulting state, a state of knowledge in the case that we are considering, a state of knowledge that yields specific examples of propositional knowledge. We can do better than mere speculation; there is, in fact, a reasonably successful empirical theory that gives substance to this proposal, namely the theory that postulates the projection principle, the binding theory, the theories of government, case, control, locality and others, along with certain parameters, as a component of the initial state. In contrast, no proposal of even minimal

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

plausibility exists to account for facts of the sort illustrated in terms of some sort of 'learning' or 'confirmation', very much as with puberty or the rigidity principle.

Certainly what I have just suggested could be correct; indeed, there is fairly good reason to suppose that something of the sort *is* correct. Since it could be correct, it cannot be that propositional knowledge must have adequate grounds or justification or warrant in anything like the sense assumed in modern epistemology. This cannot, in short, be a conceptual requirement associated with the concept 'knowing that'. The reason is that we have clear examples of such knowledge that may violate (and apparently do violate) this requirement. Therefore, it is a contingent question of fact whether specific examples of knowing-that in other domains are grounded or rather simply 'grow in the mind', as appears to be so with respect to much of linguistic knowledge.

Suppose, for example, that we observe a moving object passing behind a barrier and we know that it will emerge at such and such a point. Surely there are such examples of propositional knowledge. Is this justified knowledge, or is it a consequence of some innate system of principles perhaps triggered by experience? Or suppose that we observe a plane figure perpendicular to the line of sight and when this figure rotates until it is parallel to the line of sight, we see it as a rotating plane figure. That is, we know that it is a rigid rotating plane figure and not (as it might have been) a plane figure shrinking to a line. Is this a case of justified knowledge, or is it a consequence of something like the rigidity principle? These are questions of fact, and in these and many similar cases it seems likely that the fact is that the knowledge is not grounded. If some organism had different built-in principles, distinct from the binding theory or the rigidity principle and so on, it would simply have different beliefs and even a different domain of knowledge, including propositional knowledge, on the same evidence. It seems to me reasonable to speculate that a substantial part of our knowledge about language, about the behaviour of objects in three-dimensional space, about other people, and the like, is knowledge that grows in the mind on the basis of a system of innate principles, rather than knowledge that is grounded in experience; though the latter category too exists no doubt, in domains where built-in structure is inadequate: are all ravens black: are there free quarks? etc. At least, these are factual questions, not resolvable on *a priori* grounds. If so, then our general approach to questions of the nature of knowledge and belief requires considerable rethinking, in a fundamental way.

Language and grammar; rules and principles

I shall conclude with a few observations on the ways in which ideas about language have evolved within the past several decades, at least in the work I have been considering here. I think that there have been two significant shifts of focus with regard to the way in which the major problems are conceived. The first is the shift of focus from language to grammar, which took place about 25 years ago. For structural linguistics and its predecessors, the object of inquiry was language, and correspondingly, the closest analogue to u.g. was the theory of general properties of many or all languages. The approach that I have been discussing took a radically different stance. The focus of inquiry is grammar; language is a derivative and possibly uninteresting notion. The shift of focus was sometimes obscured in introductory expositions or in mathematical linguistics, but it was clear in the earliest work on generative grammar. I think that this shift was quite proper. Contrary to what is commonly believed, the notion 'language', if it is even coherent, is at a much higher level of abstraction from actual

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

mechanisms than grammar. Grammars exist in the world, as components of steady states attained. As for languages, one may perhaps think of them as determined by grammars, or in some other way, but in any event they are clearly at a further remove from real mechanisms of the brain than the grammars represented in these mechanisms. Correspondingly, the theory of u.g. is not the study of general properties of language, but rather u.g. is a postulated component of genetic endowment, on a par with the properties that determine that the embryo will grow arms rather than wings, use the rigidity principle, undergo puberty, etc. – all, of course, under appropriate external conditions. Once this change of focus is adopted, this part of linguistics becomes part of psychology, and ultimately biology. It is a subdomain delimited by the uninteresting characteristic that its practitioners, for the moment, happen to rely primarily on certain kinds of evidence, such as those illustrated earlier, because such evidence appears to be most useful in advancing their inquiry into properties of the initial and steady state of the language faculty. There is every reason to hope that this artificially delimited discipline will disappear as other kinds of evidence become available that bear on the questions with which it has been concerned.

The second shift of focus is more subtle, and its significance is just beginning to become clear. Early theories of generative grammar permitted a wide range of possible grammars in the interest of attaining descriptive adequacy, and the associated theory of u.g. was correspondingly limited, though far from vacuous, in explanatory power. Much of the work of the past 20 years has been devoted to restricting the class of available grammars so that explanatory adequacy is enhanced, as it becomes possible to determine why the steady state takes the specific form that it does. The general approach has been to limit the variety of possible rules while formulating general principles, with a few parameters, that restrict the ways in which rules may apply. As this work has progressed, the focus of attention has gradually shifted from rule systems to systems of principles such as the binding theory and the projection principle. In fact, current theories of transformational generative grammar are so restrictive that they permit only a finite number of grammars in principle, apart from the lexicon. Much of the debate over the character of rule systems proves to be near vacuous, e.g. the question of whether a certain rule is a transformational rule or a rule of interpretation with exactly the specific properties of transformational rules. I say 'near vacuous' because there may be empirical differences, but if they exist, and they may, they are rather subtle. In contrast, questions concerning the subsystems of principles often have substantial empirical content.

One consequence of this second shift of focus is that most of the results of mathematical linguistics, which in any event have been seriously misinterpreted, become empirically virtually or completely empty, since they deal with properties of infinite classes of grammars. The same is true of the mathematical theory of learnability in so far as its results depend on the presumed infinity of the class of grammars. The question of generative capacity, for example, has little meaning for finite sets of grammars. The specific detailed structure of the elements of these sets now has an overwhelming effect on any results that might be obtained in studies of generative capacity or learnability. Similar remarks hold for the theory of parsing, though for somewhat different reasons, as has been pointed out by Berwick & Weinberg (1981). The mathematical results concern parsability in the limit, as sentences grow longer. But any empirically meaningful results will deal with parsability of 'short sentences', whether we are considering parsing or functional explanation for language evolution. Thus, the standard result for parsing of context-free languages shows that parsing time is proportional to the square of the size of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the grammar and the cube of the length of the input sentence. The first factor, however large, is fixed, and disappearss in the limit as sentences become longer. But for 'short sentences' of the sort that are actually used, the first factor may be overwhelming and the second insignificant. The real empirical content of existing results, then, may well be that grammars are preferred if they are not too complex in their rule structure. If parsability is a factor in language evolution, we would expect it to prefer 'short grammars' – such as transformational grammars based on the projection principle or the binding theory – and to care very little about parsability in the limit or even parsability for 'long sentences' of a sort rarely encountered in ordinary life.

Such principles as the projection principle make strong empirical predictions, and have a large effect on simplifying grammar and probably on the proper formulation of questions of parsing and learnability. It also becomes very important to determine in just what respects parametric variation is permitted. In a theory of u.g. with a fairly rich structure, a small change in a single parameter may lead to a substantial difference in the resulting 'language', as effects proliferate through the system. And a few such changes may yield languages that look very different from one another, though they are basically cast in the same mould. For similar reasons, species may seem highly diverse as a result of modifications of regulatory circuits and the like in a common biochemical system. These are the kinds of result that we hope to obtain in the study of language, or better, the study of grammar, the real object. Qualitatively speaking, languages appear to be highly diverse, yet there is strong reason to believe that the initial state is highly restrictive, so that an intricate system of knowledge, with quite subtle and specific properties, is acquired on fragmentary evidence. A theory of principles and parameters with simple subtheories and a fairly rich structure of interactions appears to be what is required to provide the actual observed results. Such theories are now, for the first time, becoming available. This is a new development, and one that I believe to be very exciting.

References (Chomsky)

Berwick, R. & Weinberg, A. 1981 Parsing efficiency, computational complexity, and the evaluation of grammatical theories. M.I.T. mimeograph.

Chomsky, N. 1977 On *wh*-movement. In *Formal syntax* ed. P. Culicover, T. Wasow & A. Akmajian), pp. 71–132. London: Academic Press.

Chomsky, N. 1980 *Rules and representations*. New York: Columbia University Press.

Chomsky, N. 1981*a* Principles and parameters in syntactic theory. In *Explanation in linguistics* (ed. N. Hornstein & D. Lightfoot) London: Longmans. (In the press.)

Chomsky, N. 1981*b* *Lectures on government and binding: the Pisa Lectures*. Amsterdam: Foris.

Higginbotham, J. 1980 Pronouns and bound variables. *Ling. Inquiry* **11**, 679–708.

Higginbotham, J. 1981 Reciprocal interpretation. *J. Ling. Res.* **1**, 97–117.

Marr, D. & Nishihara, K. 1978 Visual information processing: artificial intelligence and the sensorium of sight. *Technol. Rev.* **81**, 2–23.

Marshall, J. 1980 The new organology. *Behav. Brain Sci.* **3**, 23–25.

Rizzi, L. 1980 Negation, *wh*-movement and the null subject parameter. Università della Calabria – M.I.T. mimeograph.

Ullman, S. 1979 *The interpretation of visual motion*. Cambridge: M.I.T. Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES