Language Acquisition: Schemas Replace Universal Grammar[1]

Michael A. Arbib
Departments of Computer Science, Neurobiology, Physiology,
Biomedical Engineering, Electrical Engineering, and Psychology
University of Southern California
Los Angeles, CA 90089-0782, USA

and

Jane C. Hill
Department of Computer Science
Smith College
Northampton , MA 01063, USA

(Arbib,M.A. and Hill, J.C., 1988, Language Acquisition: Schemas Replace Universal Grammar, in *Explaining Language Universals*, J.A. Hawkins, (Ed.) Basil Blackwell, pp. 56-72.)

Language Acquisition: Schemas Replace Universal Grammar

Abstract

Chomskians hold that what makes it possible for the child to acquire its language in a few years on the basis of degenerate and inadequate data is the existence of an innate universal grammar. By contrast, we model language acquisition in terms of a dynamic process involving multitudinous changes in the child's stock of schemas with continuing experience. Our model demystifies the unimportance of negative data by providing a theory of the way in which positive evidence is exploited which is richer than that offered by generative theories. We suggest that universals are to be seen not as guiding the process of language acquisition, but rather as being descriptive of regularities that arise from the intricate interactions of a multitude of schemas acting and changing as the child comes to better and

1 **

1. Brain Mechanisms: Innate Capacities do not Preclude Learning

##GAP## that respect (Cowey 1981, Zeki 1984) — so can there be no argument that, as evidenced by data on lesions of Broca's and Wernicke's areas *inter alia*, the human brain is genetically specified with a network of mechanisms that make language possible. But to say that the human capacity for language is embodied in such structures does not imply that language acquisition is based on universal grammar rather than, say, the possession of innate mechanisms to relate word perception to visual perception, or to produce sentences using mechanisms evolved from those producing other types of coordinated, skillful movements. It is in the light of such considerations that we have suggested elsewhere (Arbib and Caplan 1979, Arbib 1982) that language principles are not to be understood in isolation from sensorimotor processes. In this connection, it is instructive to quote the remarks of Arbib, Caplan and Marshall (1982, pp.430-431) on the paper by Galaburda (1982) in their volume:

> ...Galaburda's evolutionary perspective on comparative anatomy emphasizes the division of cortical areas into discrete patches with distinct input and output pathways and attendant, distinctive, staining properties. In applying neuroanatomical techniques to human autopsy material, Galaburda finds an unusually high accumulation when staining with lipofucsin granules in a number of areas that are linked (through lesion analysis and neurophysiological testing) to language function. The staining selects Area 44 from the classical Broca's area, and Tpt from the Wernicke region. We thus have what appears to be the first *cellular* marker for the language areas. Moreover, in the opercular region, Area 44 alone receives direct projections from posterior auditory fields, and Area 44 and Tpt are neatly connected by fibers coursing the arcuate fasciculus. We are reminded of Geschwind's hypothesis that such a connection is an important part of the substrate for language organization in the brain. But there is one catch: The projections shown in the last sentence but one were shown in anatomical studies of the rhesus monkey. We are forced once again to place neurolinguistics in a broader neuropsychological perspective in which we can come to understand the role of these not-quite-language areas in rhesus, and thus better understand the evolutionary pedigree of our own language abilities.

Chomsky argues (Beckwith and Rispoli 1986) that we do not learn to grow arms and that we also do not learn to have language in any very interesting sense and then asserts that there are not going to be principles of learning any more than there will be principles of growth — the organs become what they do become because of genetic instructions that give them particular directions and because of the way in which the intrinsic structure relates to the environmental context which is all preprogrammed. We leave it to the embryologists to defend themselves against the slur that there are no principles of growth. Here, we want to look briefly at recent results which show that the genetic program for brain growth is open to experience in a way that reverses the thrust of Chomky's argument — it is not that the growth of language is as fixed and preprogrammed as the growth of a hand; rather it is that the theory of growth is itself beginning to develop so that it can address the subtleties of neural development in a way which may begin to make contact with our understanding of *real* learning. Hubel and Wiesel 1965 showed that cats raised with an artificial squint would "lose" stereopsis, while Hirsch and Spinelli 1970 and Blakemore and Cooper 1970 showed that cats raised in an impoverished environment would "lose" part of their normal complement of "edge detectors". However, Hirsch and Spinelli went further, showing that cats could be so trained that visual cortex neurons would be specified for "new" features not present in the normal animal. Moreover, Spinelli and Jensen 1979 have shown that the allocation of cells to different subsets of the sensory world can be modified on the basis of early experience. Fregnac and Imbert 1978 have shown that cells of visual cortex come in three varieties — totally pre-wired, biased, and totally uncommitted. The theoretical models of these phenomena by such authors as von der Malsburg 1973, Amari 1980 and Bienenstock et al. 1982 make it clear that we have a situation in which innate structure provides the basis for, rather than precluding the operation of, powerful learning mechanisms.

Universal grammar is exciting as a ##*GAP*## escription of planetary orbits justifies a theory of dynamics that holds that planets determine their trajectories by setting the major and minor axes of an ellipse. Each language has idiosyncracies of syntax that fill far more pages of the grammar books than do those general principles subsumed by "parameter settings," and learning grammar is a very small part of learning a language. The child must learn to segment the sound stream and master the idiosyncracies of the morpho-phonology of his language; to this must be added the learning of a huge vocabulary as well as a large stock of idioms, phrases and metaphors. By the time we have found explanations for the ability to learn all these, the mechanisms thus uncovered may have obviated the need for a set of parameterized universal principles. We do not to dismiss the existence of grammaticality judgements but we do claim that by examining computational models of language acquisition we may discover routes to language competence, including grammaticality judgements, that do not require a universal grammar, and such routes may in fact be discovered by focussing on language performance.

Another way of characterizing Chomsky's approach is that it proceeds *backwards* from a characterization of adult grammar to see how the child might arrive at this characterization. Our approach, by contrast, is to work *forward* from the evidence that the child provides toward some characterization of the adult language. We have already stressed that the richness of the world's languages exhibits far greater variation than is captured by the addition of parameter setting to a fixed stock of universals. We view the acquisition of language in a larger cognitive sense than does Chomsky. Our neo-Piagetian approach is to view the child, motivated by an innate desire to communicate, as actively constructing language, aided by innate cognitive schemas and mediated by the perceptive apparatus through which all humans perceive the world. To motivate our use of the term neo-Piagetian, consider that, in *Language and Learning: The Debate between Chomsky and Piaget* (Piattelli-Palmarini, 1980), Chomsky's rejection of Piaget is based more on disdain for Piaget's lack of formal precision than on any reasoned critique of the body of data the Piagetians have accumulated or the way in which Piaget's informal concepts address them. Certainly, we find Piaget at his weakest when he tries to force his rich observations into the Procrustean bed of his *groupements* ; while his description of mental development, though stimulating, lacks specific mechanisms and is overly complacent in its trust in the unfolding of stages of ever-greater sophistication (see Arbib 1987 for a critique based on the roles of instruction and historical contigency in the acquisition of concepts of logic and mathematics). However, we shall suggest below that one may provide models of language acquisition that are informed by a "computational neo-Piagetian" view of construction rather than by an appeal to innate principles of universal grammar, in that they combine a Piagetian attention to the child's "unfolding" of cognitive structure with the rigour of a computational model.

Chomsky's position is based on the view that language is too complex to be learned in the sense that one learns mathematics, or learns to play chess. Yet all normal children acquire a native language by the time that they are around five years old. Moreover, every child produces myriad original sentences. By what processes can these facts be explained? In 1965, Chomsky posited a need for a set of evaluation criteria for choosing between the presumably infinitely many hypotheses about language structure that might be compatible with the linguistic input data. The need for these evaluation criteria has since been obviated by his assumption of a set of language universals and a set of parameters that narrow the hypothesis space. Chomsky defined a universal grammar to be a system of principles which characterize the class of biologically possible grammars. Emphasizing the biological foundations of language, Chomsky likens the "growing" of language to the growing of any other organ of the body (Beckwith and Rispoli 1986). The child will hear the language of his or her environment and the discovery, for example, that this language uses subject-object-verb word order, might act as a trigger for a set of related assumptions such as that the language uses a case system. Universal grammar then has highly restricted options and a few parameteric variations.[2] Chomsky sets the debate in terms of the "setting up" of adult language rather than in terms of the dynamic changes that the child's language undergoes during the process of acquisition.

---

[2] But are the options highly restricted, and are the variations indeed few? The problems with this approach are well illustrated in the companion article by Hoekstra and Kooij (this volume). They give "(1d)*His proof of the theorem wrong" as an example of a construction of a type impossible in all languages — yet it is in fact correct in Chinese (on omitting "the"). Again, they note that the use of a parameter is required to "save" the subjacency principle for Italian, but then note that even this is not enough to make the principle of universal application. One may thus be excused for favouring a mechanism which can generate schemas to embody experienced regularities over a theory of language acquisition which is little specified beyond positing the necessity for a baroquely epicyclic data base parametrized to express variations in the thousands of human languages that the child will never learn.

Our book *From Schema Theory to Language* (Arbib, Conklin and Hill 1987) looks at language from a perspective in which performance, rather than competence or syntax, takes center stage, and builds on insights from brain theory and artificial intelligence (AI) to sketch the evolution of schema theory as it models three phenomena of language performance: language understanding by aphasics; language learning by a two-year-old (to be discussed in some detail below); and scene description. It is the second of these that will focus our discussion here. Hill (1983; see also Arbib, Conklin and Hill 1987, Part III) found that the language of a two-year old changes week by week, and offered a computational model of the learning mechanism which could underlie such changes. The model is a repetition-and-response model which explains both how the two-year-old child that she studied responds to adult utterances, usually with a truncated form of that utterance, and how the child's linguistic and conceptual structures may change with each such "repetition." It is important to note that the repetitions differ markedly from adult syntax, but do have a coherent structure whose unfolding the model addresses. *Every* adult utterance can serve to modify the child's evolving representation, and thus the model is not vitiated by Chomsky's observation that children receive little in the way of *explicit* syntactic error feedback and seem resistant to what they do receive. As we shall see below, the model requires no negative evidence, and yet successfully hypothesizes a process of dynamic change of an evolving set of word classes and grammatical templates, rather than the all-or-none acquisition of adult grammaticality. The learning process is highly dynamic, and what is learned depends upon what has been learned before, so that the same adult input data presented at different times to the model result in different patterns of learning. Thus, for the learning processes posited here, it is far from true that there is a poverty of data or that the child had no relevant experience. Hearing hundreds of sentences a day and using billions of neurons to do so, surely the wonder is that the child takes so long! However, our real claim is that the child is not so much trying to model adult criteria for syntactic well-formedness as coming to interact with, perceive, represent, and communicate about its world in ever more complex ways. Chomsky does not help us unravel these intertwined processes of construction.

We agree with Chomsky that some machinery has to be innate in the brain — but the question is whether it involves learning-principles governing a rich set of interacting subsystems, or whether it involves setting a few parameters. As support for basing language acquisition on mechanisms incorporating universal grammar, Chomsky argues (Beckwith and Rispoli 1986) that to learn whether your language is "head first" or "head last," English or Japanese, it is enough to hear three word sentences like "John saw Bill" or "John Bill saw." If you hear one, "John saw Bill," you have a "head first" language — so it is just a matter of setting parameters in a very highly constrained situation. However, from our cognitive viewpoint, this begs a multitude of questions. We must first ask "How does the child learn to recognize John and "John," Bill and "Bill," and recognize what action "saw" denotes, and who saw whom?" "How does a complex perceptual structure gets mapped into a simpler structure of words?" "How can the child recognize the order of the words in an utterance that its hears?" All this reinforces our point that a psychology of language must explain so much of language learning that is not explained by universal grammar that it is moot whether an adequate theory of such learning would leave any lacunae that universal grammar must necessarily fill. Once one has the mechanisms for all these processes, one can then address the question whether the discoveries attendant upon the "head last" vs. "head first" distinction must be learned by parameter setting or are "automatically" given by the very processes necessary to observe it.

2. Computational Models and the Importance of Errors

The approach to language acquisition embodied in both models to be described here is to work forward from the evidence that the child provides toward some characterization of the adult language. We ask not only under what conditions a language can be learned by a computational model, but also does the model learn the language in the same way that the child does? To define "in the same way," particular attention must be paid to the "errors" that children make since child speech that differs from adult speech yields clues concerning the processess that the child uses to understand and produce speech. An equally important clue to the child's processes are those errors which the child typically does not make. One frequently noted phenomenon is the over-generalization of the plural of nouns and of the past tense of verbs in English. Such phenomena need explanation. Rumelhart and McClelland's model (see Section 5) deals specifically with the learning of past-tense forms in English. Hill's model (to which we now turn) learns these forms within the context of learning word classes and syntactic constructs. The model learns to understand and generate sentences of ever greater complexity as does the child. Moreover, to provide a satisfying explanation of the course of language acquisition the model must make the same kinds of errors that the child makes, and must eventually correct the errors after further learning has occurred. As we shall see, both models proceed without the need for negative evidence. It is gratifying to see that the basic paradigm which was initially used in Hill's model to learn word classes and a simple template grammar could be quite naturally extended to the learning of past-tense forms of verbs in English. We offer these two models as examples of the different sort of answers which computational models may suggest to traditional questions, thanks to an approach to language acquisition based on a set of dynamic processes rather than a set of static rules.

3. The Hill Model

Figure 1 shows the components of Hill's model. The model takes as its input adult sentences together with indications (provided by the modeler, where relevant) of the physical context in which the sentences are uttered. Output from the model is a representation of child-like sentences repeating or responding to the adult input in accordance with the current state of the model's linguisitic capacity. The child's knowledge is represented by dynamic data structures encoding the child's lexicon, the child's grammar, the conceptual knowledge of the child, and the physical context of the dialogue. The model is given a basic lexicon and a set of concepts with a

Figure 1. Basic components of the Hill model of language acquisition.

mapping between the two. No assumptions have been made about the ultimate form of the adult grammar nor about what must be built into the model, but a precise account is kept of the knowledge and processes found necessary even for this elementary level of language understanding and production. Processes attend to the adult input and use rules of salience to focus on examples within the adult data which are used as the basis for language growth. The input data is in no way especially coded for the model, but is generally taken from language acquisition corpora of adult-child dialogue. The world knowledge is encoded in a semantic net as are the grammar templates and the lexicon. The model uses its language experience (i.e., the processing of the input sentences) to build a grammar which is at first a flat template grammar but which eventually evolves into a procedural grammar which may be described, if one chooses, by a set of recursive context-free phrase structure rules. The model embodies 5 assumptions:

1. The child has schemas for and talks about relations.
2. The child has schemas for and employs word order in his utterances.
3. The child employs processes of concatenation and deletion.
4. The child forms classes of concepts and classes of words.
5. The classifying process causes successive reorganizations of the information stored.

Thus we do *not* assume that the lexical classes of adult grammar are innate. Rather, we posit a process of *classification through word use* whereby words that are used in similar ways come to be assigned to the same class, thus extending from members of the class to further members of the class certain patterns (templates) of word use. The initial grammar is given by a set of templates, consisting of a "relation" and a "slot", which is free of any characterization of the adult grammar which will emerge but is not yet present. Hill observed a brief stage in which the child concatenated two-word templates with a common word, as in

little bear baby bear

but these soon give way to such three-word templates as

little baby bear.

The four-word utterances with repeated lexical items occurred in such a brief interval that Hill hypothesized that the three-word utterances were arrived at by (1) concatenating the two templates "little bear" and "baby bear", and (2) collapsing the concatenated relations into a single three-word utterance by deleting the first occurrence of the repeated word. Some evidence that the concatenation best captures the semantics of such three-word utterances in the young child is given by the finding of Matthei 1979 that the child interprets "the second green ball" as "the ball which is second and green" - in fact, several children, when presented with an array in which the second ball was not green, actually rearranged the balls in order to make the situation conform to their interpretation of the words.

From an adult sentence such as "daddy gave the toy to the boy" the model might initially respond with a single word such as toy. A subsequent presentation of the same sentence might cause the model to acquire a template for gave toy where gave would be classified as a relation-word and toy as a slot-filler. Yet another presentation of the sentence might cause the model to learn the template Daddy gave where Daddy was a slot-filler, and eventually the template (slot1 gave slot2) might be learned for Daddy gave toy. What is learned in each presentation of the input depends upon the language experience of the model and what has been learned so far. Thus learning is highly dynamic in that each time the same input sentence is presented to the model a different set of grammar rules and additional lexical class information may be learned.

No information is given the model about word classes, but hearing sentences such as "mommy gave the toy", "John gave the book", "Sue gave the puzzle", would eventually cause the model to put toy, book, and puzzle all together in a word class of words which stand for possible objects of the relation-word gave. Note that it would not matter if the input sentences were far more complex than those used here for illustration. If the model is focussing on the word gave then a sentence such as "Mommy gave the toy to Sue while she went into the store to buy groceries" would have just the same effect as the short sentences used above. By this process word classes are derived from the model's ability to produce language. The process results in a multiplicity of overlapping and intersecting word classes. The model requires schemas for word classification and template classification in order to grow, but the actual classes remain flexible. Processes of generalization eventually also permit the classifying of relation-words which might permit, for example, giving and bringing to be relation-words that could be classed together as words which have similar syntactic properties.

Successive reorganizations of the grammar and the lexicon occur as learning takes place. This process of gradual broadening of word classes and grammatical rules from applying to specific examplars to applying to sets of specific examplars and thence to more general categories has been defended by Kuczaj 1982 and Maratsos and Chalkely 1980. In this fashion the model suggests one way in which language based initially on cognitive knowledge can grow into a syntactic system which will be increasingly independent of its semantic and cognitive foundations. It is important to note that the although the rules embedded in processes within the model are simple, their interaction is complex enough to necessitate the use of a computer model.

4. Dynamic Rule Schemas and the Use of Weighted Hypotheses

Why does the child not end up with an overly generalized grammar or lexicon? There is much discussion in the literature concerning the kinds of generalizations and over-generalizations that children make (e.g., Brown 1973 and deVilliers and deVilliers 1978). We believe that it is important to focus on the errors that children make because of the insights which they yield concerning the processes that the child employs in language acquisition. Bowerman 1974 states this position very clearly. A study by Bybee and Slobin 1982 presents a careful examination of the acquisition of irregular past-tense forms of verbs in English. If, however, we permit no overt and specific correction of the child's errors, then how shall we explain why errors of over-generalization do not persist into adult speech?

It is especially interesting to explore the use of verbs in English in the developing language of the child since learning English is intimately tied to the learning of verbs. DeVilliers 1985 has found evidence that input language has a significant impact on the child's developing language with respect to verbs. The mother's use of verbs is a high predictor of the child's use — it is not the frequency of the mother's use, but rather the variety of verb forms in the mother's use which is significant. This is interpreted to mean that the child is monitoring the input for clues about the prototypicality of forms of individual verbs. Wide differences in the use of verbs between subjects were found in the samples considered in her study. Verbs with a variety of heard uses were used with greater confidence by the child even in unheard contexts. Hill's model simulates this monitoring process. DeVillier's analysis did not address the issue of over-generalization but it does lend credence to our processes which rely on the information gleaned from the input by focussing on different constructs at different times for the learning of forms. Thus whether one concedes that the input has an impact on the language learning of the child depends on whether one monitors the course of development over fine time slices, or whether one simply looks at the end product (adult language).

Consider the verb break. It is an empirical fact that children at the earliest stage of language acquisition typically learn the word broke and seem to use it correctly. One may assume that such forms have been learned by rote. Then at a subsequent stage of development the child will start to use the word breaked. Is this because the child has formed a general schema for forming the past tense of verbs? Eventually of course children learn that break is an irregular verb and does not obey the general rule in the forming of its past tense. But the puzzle is that for a period of time, sometimes for years, both forms exist in the child's vocabulary. How can this period of imbalance between the erroneous and the correct forms be explained? It cannot be explained if the language mechanisms are expressed in terms of explicit rules which the child either does or does not know.

One answer to this question is proferred by our computational model which attaches a weight representing a degree of confidence is associated with each hypothesis about word forms or grammar rules. (The use of weights to direct learning in computational models is by no means new; see, e.g., Kelley 1967.) Each time an adult sentence is presented to the model, it is searched for possible instantiations of the available set of hypotheses; and the child's output also represents an instantiation of a particular hypothesis. The crucial point is that this process of instantiation involves a competition which depends on the current weights associated with the various hypotheses, and that these weights are themselves changed in the process. The weight associated with a hypothesis is increased each time that hypothesis is instantiated in the adult speech input. The weight is similarly increased, but to a lesser degree, each time a hypothesis is employed in the child-like output. In this way more frequently matched constructs are preferred over (given more weight than) less frequently matched constructs. Hypotheses must be reinforced to survive. If new hypotheses, however, are to start with very low weights they will have trouble "catching up" with earlier hypotheses. For this reason, separate recency values are employed whose function it is to cause more recent hypotheses to be favored for testing.

We illustrate the use of weights in the learning of past tense forms. In order to observe the correlation between past-tense and -ed endings, the model must be given a representation of time-past in its cognitive knowledge, and the ability to identify action verbs in its lexicon. The model forms past-tense entries in its lexicon for all action verbs simply by adding -ed endings. Each of these forms is initially given a modest confidence factor. The model then proceeds to modify the confidence factors of the past-tense forms depending on its language experience. The confidence factor of a form is incremented is added each time that the model recognizes a past-tense form in the adult input; a smaller increment is added each time that the model produces a past-tense form. This general scheme has the advantage that for a period of time when confidence factors are approximately in balance, two or more constructs can co-exist, as for example in the case of the past-tense over-generalization breaked and the correct form broke. Since the choice of past-tense form depends upon the history of the model, no *a priori* conclusions can be drawn about the specific past-tense forms which are learned, but depending upon the input data, the model (1) may keep an erroneous -ed ending, (2) may proceed through a period of instability in which the output vacillates between an erroneous -ed ending and the correct irregular form, or (3) may discard the erroneous form and replace it by the irregular form (see Hill 1986 for further description of the model with regards to past-tense forms).

A paradigm such as ours may thus be sensitive to the input data and may exhibit varied behavior without the need for negative evidence. Dynamic rule schemas and confidence factors has been used to model the phenomena of generalization, over-generalization, and subsequent correction of over-generalized forms. Thus we need not talk of rules or individual cases which have been learned or have not yet been learned but rather of a continuum in which rule procedures are either strong or weak.

Other issues which the same model explores are what variation occurs in the model as specific constraints are built-in or omitted, how input filters can focus on different aspects of the input data over time, and how variation in meaning representation and sets of semantic features can affect the learning process. Recently, Hill has begun to explore the effects of encoding phonological data in the input to the model. A pilot study conducted with Ann Peters (building on the studies by Peters 1985a,b and Wilson & Peters 1984) has shown that in order to model her corpus of data collected from a blind child, the inclusion of primary and secondary stress and of intonational boundaries is of great importance.

5. The Rumelhart and McClelland Model

To further advance our argument, we now discuss Rumelhart and McClelland's 1986 model ofhow children may learn the past-tense forms of verbs in English. Since these past-tense endings exhibit a highly idiosyncratic structure peculiar to English, a model of their acquisition can make no appeal to universal grammar. The point again is that, once we have modeled the "non-Chomskian" processes involved in this phase of language acquisition, we have increased the inductive evidence for our schema-based approach to language acquisition in general. The present model is at a finer grain than Hill's model, approaching the learning of past-tense verb forms from the phonological level. It is an example of a "connectionist" model, in that knowledge is not encoded in a small set of explicit rules, but is embedded in the connections between a large number of simple processing units (neuron-like, but not to be confused with actual neurons in the child's brain). The model interactively activates subsections of the network of these simple units. As in a neural network, each unit sums its inputs, excitatory and inhibitory, from other processing units to determine its output which can then affect other processing units. The continued interaction of these excitatory and inhibitory effects causes the network to converge on a decision about a hypothesis — through distributed interactions, not through the sovereignty of any single rule.

The Rumelhart and McClelland model learned the past tense of some 420 verbs in English, some regular and some irregular. The model explains the period of instability between correct and incorrect forms, and moreover the model output evidences a rough correlation between the difficulty of learning particular forms and the observations of Bybee and Slobin (1982) concerning the course of learning in the child. How difficult a word form is to learn in the model depends upon the corpus as a whole. What is crucial here, and in agreement with the Hill model, is that (a) cognitive science must address the time course of mental development, not just adult competence; and (b) the model makes no appeal to explicit representation of a general rule. Rather, the decentralized interaction of many components, representing different verbs, yields a coordination of their behavior which is describable by a rule, but which in no way is the expression of any such rule, innate or otherwise. We believe that the development of models such as these will have a large impact on future work in language acquisition.

The occurence of errors of over-generalization and their subsequent correction is totally unexplained by any of Chomsky's theories. Such phenonema are not deemed important by Chomsky and if noticed at all are dismissed as belonging to "pre-language" (cf. Wanner and Gleitman 1982 on the "tadpole/frog" hypothesis). Chomsky speaks of the child making mistakes because he "simply doesn't yet know how a parameter is set" (Beckwith and Rispoli 1986). But, in fact, few of us who have not read the Chomskian canon even know that the parameters exists (if indeed they do). If we use the more careful phrasing "the child's behavior does not yet exhibit the regularities describable by the setting of the parameter", we leave open the hypothesis that such regularities are, as suggested above, descriptive rather than causal.

6. And yet, there are rules

Since the Hill model deals only with very early stages of language acquisition it is open to the criticism that the subtleties of language which are yet to be learned by the model are precisely the areas of language acqusition addressed by universal grammar. We do not believe this to be the case, but we cannot yet offer a full-fledged computational model of language acquisition. We offer this model as an illustration of the kinds of processes which may enable language acquisition to be bootstrapped by gleaning example templates from adult input, given innate schemas of the kind we have posited thus far. Our weighting schemes provide an answer to the problem of the lack of negative evidence, and our focussing mechanisms suggest how the input data may be used. Our use of the distributional data encoded in the child's own production data, based initially on cognitive knowledge, suggests a way in which word classes may come to be formed. At the very least the model illustrates a manner in which all the input data may be processed, yet with only selected portions of that data focussed upon at different times. By contrast, the theory of universal grammar is based on no attempt to constructively assess the data on, or models of, language acquisition developed to date, and fails to address those issues of language as a medium of communication that are of concern to many cognitive scientists.

To highlight what we have shown, we close by discussing two passages from Hoekstra and Kooij's contribution to this volume, for they are, alas, typical of the "method" of those who advance universal grammar as a "model" of language acquisition. In the first passage, they downplay the argument that a string of words might be judged ungrammatical in English because the pattern to which it conforms is never induced as a possible pattern in English, noting that "this objection implies the assumption that users keep track of the patterns that they have encountered, which is a very strong assumption." Since this appears to be the basic assumption in our model, we must both defend it and undermine Hoekstra and Kooij's argument. First, note that current versions of universal grammar require that each word have a complex entry in the lexicon, and no claim is made that these entries are innate. Thus the proponent of universal grammar is committed to having the child learn a great number of patterns, for surely it is not too strong an assumption that language users keep track of the words that they have encountered. Second, just as we might not expect the user to remember every occasion on which he had heard a word and yet still to have encoded in memory a "spanning set" sufficient to establish the lexical entry and a set of usages for each word, so is our argument unweakened if we allow the Hill model to have a "spanning set" of patterns rather than a complete set of all patterns ever encountered. Third, we stress that the Hill model presents the first stages in the acquisition of patterns of ever greater abstraction and generality, so that the adult's grammaticality judgement is not based on an exhaustive search of every string of words ever encountered, but rather involves rapid access to patterns at the appropriate level of generality through word classes at that level — but word classes built up through experience (our notion of "classification through word use"), rather than given as *a priori* universals.

Later on in the same paragraph from which we have just quoted, Hoekstra and Kooij ask us to consider (2a) versus (2b):

(2) a. Where did John say that we had to get off the bus?

b. Where did John ask whether we had to get off the the bus?

Both sentences are grammatical. However, while (2a) is ambiguous between a reading in which where has matrix scope or embedded scope, (2b) can only be interpreted as a question concerning the place where John had uttered a particular question. This piece of knowledge is shared by all native speakers, but it can hardly have been established on the basis of induction, simply because there are no data from which induction could conceivably proceed.

This is breathtaking! Where is the evidence that "this piece of knowledge is shared by all native speakers"? It certainly is not shared by all two-year old native speakers. Perhaps Hoekstra and Kooij mean "all *adult* native speakers." But if so, and even if they are correct, where is the explanation for the transition that occurs in the individual child from ignorance to knowledge? By parameter setting? But where is the explicit description of what the child's language would look like before and after the parameter is set? And where is the corpus collected from a large number of children showing the magical moment at which the crucial datum sets the parameter? We would assert that there is no one crucial datum. Rather, the child is exposed to a vast array of data which include questions whose patterns are reflected by (2a) and (2b). The very patterns that enable the child to master the words "ask" and "say," "that" and "whether" are the same patterns which give the child "the data from which induction could .. proceed" whose existence Hoekstra and Kooij deny. But they deny it without *any* empirical analysis of how the language of the child changes with experience.

We have *shown* that simple patterns can evolve into complex patterns in a way which matches patterns of language acquisition in a two-year-old child. We *claim* , as a target for future research, that our model can be extended (not by the formation of word patterns alone, but [cf. Figure 1] through the continuing interaction of the lexical space, grammar space, and cognitive space of the child) to cover such phenomena as the distinction between (2a) and (2b). Yes, an adult (and one trained in linguistic terminology at that) can learnedly discern patterns in (2a) that can be distinguished as matrix scope vs. embedded scope, but this adds no weight at all to the claim that such knowledge is embedded in an innate universal grammar, and that without "knowing" matrix scope vs. embedded scope innately, the child could not acquire the ability to distinguish (2a) from (2b). In our society, children initially learn language through using it without any necessary reflection upon its patterns. Eventually, the child does come to reflect with pleasure on these patterns, as well as to experience, with less pleasure, the explicit presentation of grammatical rules in the classroom. But this does not concede an explanatory role for universal grammar in language acquisition. We argue that the rules are structures whose acquisition is made possible by the prior acquisition of language, not innate structures that make the acquisition of language possible.

References

S.Amari (1980): Topographic organization of nerve fields, *Bull. Math. Biol.* 42: 339-364.

M.A. Arbib (1982): Perceptual-motor processes and the neural basis of language, in *Neural Models of Language Processes* (M.A.Arbib, D.Caplan and J.C.Marshall, Eds.) Academic Press, pp. 531-551.

M.A.Arbib (1986): Minds, modules and schemas: A response to Chomsky, *New Ideas in Psychology* 4:203-210.

M.A.Arbib (1987): A Piagetian perspective on the construction of logic, *Synthese* (in press)

M.A.Arbib and D.Caplan (1979): Neurolinguistics must be computational, *Behav. Brain Scis.* 2:449-483.

M.A. Arbib, D.Caplan and J.C.Marshall, Eds. (1982): *Neural Models of Language Processes,* Academic Press.

M.A.Arbib, E.J.Conklin and J.C. Hill (1987): *From Schema Theory to Language*, Oxford University Press.

R. Beckwith and M.Rispoli (1986): Aspects of a theory of mind: An interview with Noam Chomsky, *New Ideas in Psychology* 4:187-202.

E.L.Bienenstock, L.N.Cooper and P.W.Munro (1982): Theory for the development of neuron selectivity; orientation specificity and binocular interaction in visual cortex, *J. Neuroscience*, 2:32-48.

C.Blakemore and G.Cooper(1970): Development of the brain depends on the visual environment, *Nature* 228:477-478.

M. Bowerman (1974): Learning the structure of causative verbs: A study in the relationship of cognitive, semantic, and syntactic development, in *Papers and Reports on Child Language Development,* no. 8 (E. Clark, ed.) Stanford University Committee on Linguistic, pp. 142-178.

R. Brown (1973): *A first language: the early stages* , Harvard University Press.

J. Bybee and D. Slobin (1982): Rules and schemas in the development and use of the English past tense, *Language,* 58:265-289.

N. Chomsky (1977): *Aspects of the Theory of Syntax* , The MIT Press.

A. Cowey (1981): Why are there so many visual areas? in *The Organization of the Cerebral Cortex* (F.O.Schmitt, F.G.Worden and F.Dennis, Eds.) The MIT Press, pp.395-413.

J. deVilliers (1985): Learning how to use verbs: lexical coding and the influence of the input, *Journal of Child Language* 12:587-595.

J. deVilliers and P. deVilliers (1979): *Early language* , Harvard University Press.

Y.Fregnac and M.Imbert (1978): Early development of visual cortical cells in normal and dark-reared kittens: Relationship between orientation selectivity and ocular dominance, *J. Physiol. London* , 278:27-44.

A.M.Galaburda (1982): Histology, Architectonics, and Assymetry of Language Areas, in *Neural Models of Language Processes* (M.A.Arbib, D.Caplan and J.C.Marshall, Eds.) Academic Press, pp. 435-445.

J.C.Hill (1983): A computational model of language acquisition in the two-year-old, *Cognition and Brain Theory* , 6:287-317.

J.C.Hill and M.A.Arbib (1984): Schemas, computation and language acquisition, *Human Development* , 27:282-296.

J.C.Hill (1986): A Computational model which addresses errors of overgeneralization and their subsequent disappearance, *Proceedings of the Eight Annual Conference of the Cognitive Science Society,* University of Massachusetts at Amherst, August.

H.Hirsch and D.N.Spinelli(1970): Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats, *Science* 168:869-871.

T.Hoekstra and J.G.Kooij (this volume): The innateness hypothesis, in *Explaining Language Universals*, (John A. Hawkins, Ed.), Basil Blackwell.

D.H.Hubel and T.N.Wiesel (1965): Binocular interaction in striate cortex of kittens reared with artificial squint, *J. Neurophysiol.* 28:1041-1059.

K. Kelley (1976): Early syntactic acquisition, Ph.D. Dissertation, University of California at Los Angeles, 1967, also published as Report No. P-3719, The Rand Corporation, Santa Monica, CA., November, 1967.

S.A. Kuczaj II (1982): On the nature of syntactic development, in *Language Development: Volume 1: Syntax and Semantics,* (S. Kuczaj, ed.) Lawrence Erlbaum Associates.

M. Maratsos and M. Chalkley (1980): The internal language of children's syntax: the ontogenesis and representation of syntactic categories, *Children's Language*, vol.2 (K. Nelson, ed) Gardner Press.

E.Matthei (1979): The acquisition of prenominal modifier sequences: stalking the second green ball, Ph.D. Dissertation, Dept. of Linguistics, University of Massachusetts at Amherst.

M. Piattelli-Palmarini, Ed. (1980): *Language and Learning: The Debate between Jean Piaget and Noam Chomsky* , Harvard University Press.

A.Peters (1985a): The role of imitation in the syntactic development of a blind child. Paper presented at the Society for Research in Child Development, Toronto, April.

A.Peters (1985b): Routines as loci for language development. Paper presented at the Boston University Conference on Child Language, October.

D.E.Rumelhart and J.L.McClelland (1986): On learning the past tense of English verbs, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* , (D.E.Rumelhart and J.L. McClelland , Eds.), A Bradford Book /The MIT Press, vol.2, pp.216-271.

D.N.Spinelli and F.E.Jensen (1979): The mirror of experience, *Science* 203:75-78.

C.von der Malsburg(1973): Self-organizing of orientation sensitive cells in the striate cortex, *Kybernetik* , 14:85-100.

E.Wanner and L.Gleitman, (1982): *Language Acquisition: the state of the art* , Cambridge University Press.

B.Wilson and A.Peters (1984): "What are you cooking on a hot?": a blind child's "violation" of "universal" constraints. Paper presented at the Boston University Conference on child language, October.

S.Zeki (1984): The specialization of function and the function of specialization in the visual cortex,in*Recent Advances in Physiology* (P.F.Baker, Ed.) Churchill Livingstone, pp.67-85.

*****

Language Acquisition: Schemas Replace Universal Grammar

Michael A. Arbib
University of Southern California
Jane C. Hill
Smith College

As background for this talk, I mention 3 recent books:

• E.G.Manes and M.A.Arbib: *Algebraic Models of Program Semantics* , Springer-Verlag, 1986.

Algebraic models for the semantics of programming languages, including such constructs as recursion, and functorial approaches to data types.

(2)

• M.A.Arbib and M.B.Hesse:*The Construction of Reality* , Cambridge University Press, 1986.

Explores a number of philosophical issues from a schema-theoretic perspective,analyzing the construction of knowledge in both individualist and social terms. It offers a view of language as inherently metaphorical.

• M.A.Arbib, E.J.Conklin and J.C.Hill: *From Schema Theory to Language* , Oxford University Press, 1987.

An account of our computational models of language performance based on the interaction and modification of schemas.

(3)

The present talk recapitulates material from Part I An Overall Perspective, and Part III Language Acquisition of this last volume, and then uses the discussion of a schema-based model of language acquisition to ground a critique of the Chomskian account of acquisition which appeals to an innate universal grammar.

Schema theory is not yet a theory in the sense of a core of time-tested definitions and theorems but is emerging as a "federation of mini-theories" from studies of schema-based theories of vision, motor control — and language.

(4)

A map or a model must omit many details to focus upon essential issues. To bridge from overall animal behaviour to detailed neural circuitry brain theory must proceed at *many different levels of analysis.*

Bottom-up analysis starts with the detailed working out of the interaction of individual neurons in explaining network properties. *Top-down analysis* achieves, I claim, a functional model of some overall behaviour through the interaction of a number of simultaneous computing agents called *schemas.*

Most succesful modelling will at most be bottom-up or top-down in its initial stages.

(5)

Schemas are functional units

• "large" enough so that an overall behaviour may be analyzed in terms of interacting schemas

• "small" enough that the constituent schemas can be implemented in anatomically and physiologically testable neural circuits, e.g., via *neural layers* as structural units.

Cooperative Computation of interacting subsystems provides the style of computation at all levels: Interaction between concurrently active schemas/regions of the brain, rather than one-way flow of information in a hierarchically organized system. Different systems each with only partial sources of information cooperate to determine the overall behaviour of the organism.

(6)

Example: The Cue Interaction Model of Depth Perception:

Both accommodation and disparity provide partial representations.These are coupled to make it more likely that the animal will obtain an accurate representation of the depth of objects.

The linkage of subsystems implies that the activity of each system encodes pooled information.

The choice of some overall animal behaviour for study is theory-laden. A "natural" unit of behaviour may not to be unitary — e.g., quite different mechanisms seem involved in a toad's depth perception for prey and for barriers.

(7)

Claims about the localization of schemas in the brain may be tested by lesion experiments.

In general we cannot expect a simple match between a structural and functional analysis:

A given schema, defined functionally, may be distributed across more than one brain region

A given brain region may be involved in many schemas.

For the original application of these ideas to neurolinguistics, see

• M.A. Arbib and D.C.Caplan: Neurolinguistics Must be Computational, Beh. Brain. Scis. 1979;

• M.A.Arbib, D.C.Caplan and J.Marshall (Editors):Neural Models of Language Processes, Academic Press,1982.

(8)

There is no single definition that encompasses all programs whether serial or parallel or concurrent; whether recursive or not; whether object-based or not. Similarly, our work on schemas to date yields no single formalism but contributes to the evolution of a theory of schemas as programs meeting the following criteria:

• They serve to represent, at least, processes/ representations for perceptual structures and distributed motor control;

• Schemas may be instantiated: E.g., given a schema that represents generic knowledge about a chair, we need several instantiations each suitably tuned to subserve our perception of several chairs;

• The programs are concurrent. Unlike serial computers, the brain can support concurrent activity of many schema instantiations.

(9)

The actions of common (human) behaviour depend on far more than current sensory stimulation. As the organism moves — making, executing, and updating plans — it must maintain an up-to-date representation of its relationship with its environment. We posit that the *model of the environment* is an active, information-seeking process built as an *assemblage of perceptual schemas* each of which roughly corresponds to a *domain of interaction.*

A *perceptual schema* embodies the process for determining whether a given domain of interaction is present in the environment.

(10)

The state of activation of an *instantiation* of a perceptual schema is a measure of the credibility of the hypothesis that what the schema represents is indeed present. Other schema parameters represent properties such as size, location, and motion of the perceived object/task situation.

Activation of perceptual schemas provides access to related *motor schemas* but does not necessarily entail execution of these schemas. *Planning* is required to determine the actual course of action. The action of the organism is controlled by a *plan* made up of *motor schemas* — akin to control systems but combinable to form *coordinated control programs* which will control the phasing in and out of various patterns of movement.

(11)

An assemblage of *instantiated* perceptual schemas provides a Short Term Memory (STM) combining an estimate of environmental state with a representation of goals and needs.

New sensory input updates the *schema assemblage* which can itself be action- dependent.

Anticipatory schemas are plans for perceptual action as well as readiness for particular kinds of sensory structure. We thus view behaviour in terms of a continuing *action-perception cycle* rather than in terms of a discrete stimulus yielding a discrete response.

(12)

The plan is updated as action affords perceptual updating of the internal model:

As action continues the current plan may continue to be executed with tuning or updating of parameters; or, because of some unexpected occurrence or completion of the current plan, some form of replanning or new planning may be required.

(13)

Behavior is as real as anatomy: A repeatable "input/output" schema is as real as a repeatable anatomical structure. A network of interacting "internal" schemas is an approximation to reality. The schemas become "more real" as their functional analysis is refined into assemblages/programs of subschemas which allow either a more subtle analysis of behavior or an improved mapping of function to neural structure.

(14)

Now: on to language acquisition:

Chomskians hold that what makes it possible for the child to acquire its language in a few years on the basis of degenerate and inadequate data is the existence of an innate universal grammar. By contrast, we model language acquisition in terms of a dynamic process involving multitudinous changes in the child's stock of schemas with continuing experience. Our model demystifies the unimportance of negative data by providing a theory of the way in which positive evidence is exploited which is richer than that offered by generative theories.

(15)

As evidenced by data on lesions of Broca's and Wernicke's areas, the human brain is genetically specified with a network of mechanisms that make language possible. But this does not imply that language acquisition is based on universal grammar rather than, say, the possession of innate mechanisms to relate word perception to visual perception, or to produce sentences using mechanisms evolved from those producing other types of coordinated, skillful movements.

(16)

Chomky argues that the growth of language is as fixed and preprogrammed as the growth of a hand — but recent results show that the genetic program for brain growth is open to experience in a way that reverses the thrust of this argument: cf. the data of Hubel and Wiesel 1965; Hirsch and Spinelli 1970; Blakemore and Cooper 1970; Spinelli and Jensen 1979; Fregnac and Imbert 1978. The theoretical models of these phenomena by such authors as von der Malsburg 1973, Amari 1980 and Bienenstock et al. 1982 make it clear that we have a situation in which innate structure provides the basis for, rather than precluding the operation of, powerful learning mechanisms.

(17)

Universal grammar is exciting as a *description* of general properties of adult syntax but we reject the claim that the setting of parameters can quickly outline the grammatical structure of all human languages, even from "unrelated families." *Even if the description were true* , this would provide no argument for Chomsky's use of "parameter setting" as a theory of language acquisition — any more than Kepler's description of planetary orbits justifies a theory of dynamics that holds that planets determine their trajectories by setting the major and minor axes of an ellipse.

Each language has idiosyncracies of syntax that fill far more pages of the grammar books than do those general principles subsumed by "parameter settings," and learning grammar is a very small part of learning a language.

(18)

The child must learn to segment the sound stream and master the idiosyncracies of the morpho-phonology of his language; to this must be added the learning of a huge vocabulary as well as a large stock of idioms, phrases and metaphors. By the time we have found explanations for the ability to learn all these, the mechanisms thus uncovered may have obviated the need for a set of parameterized universal principles.

Our neo-Piagetian approach is to view the child, motivated by an innate desire to communicate, as actively constructing language, aided by innate cognitive schemas and mediated by the perceptive apparatus through which all humans perceive the world.

(19)

We provide models of language acquisition informed by a "computational neo-Piagetian" view of construction rather than by an appeal to innate principles of universal grammar — combining a Piagetian attention to the child's "unfolding" of cognitive structure with the rigour of a computational model.

Chomsky defined a universal grammar to be a system of principles which characterize the class of biologically possible grammars. The child will hear the language of his or her environment and the discovery, for example, that this language uses subject-object-verb word order, might act as a trigger for a set of related assumptions such as that the language uses a case system.

(20)

It is claimed that universal grammar has highly restricted options and a few parameteric variations. But are the options highly restricted, and are the variations indeed few? Hoekstra and Kooij give

"(1d)*His proof of the theorem wrong"

as an example of a construction of a type impossible in all languages — yet it is in fact correct in Chinese (on omitting "the"). Again, they note that the use of a parameter is required to "save" the subjacency principle for Italian, but then note that even this is not enough to make the principle of universal application. One thus questions a theory of language acquisition little specified beyond positing an epicyclic data base parametrized to express variations in thousands of human languages that the child will never learn.

(21)

Chomsky sets the debate in terms of the "setting up" of adult language rather than in terms of the dynamic changes that the child's language undergoes during acquisition.

We claim that the child is not so much trying to model adult criteria for syntactic well-formedness as coming to interact with, perceive, represent, and communicate about its world in ever more complex ways. Our book *From Schema Theory to Language* looks at language from a perspective in which performance, rather than competence or syntax, takes center stage.

(22)

Hill 1983 found that the language of a two-year old changes week by week, and offered a computational model of the learning mechanism which could underlie such changes. The model is a repetition-and-response model which explains both how the two-year-old responds to adult utterances, usually with a truncated form of that utterance, and how the child's linguistic and conceptual structures may change with each such "repetition."

(23)

It is important to note that the repetitions differ markedly from adult syntax, but do have a coherent structure whose unfolding the model addresses. Every adult utterance can serve to modify the child's evolving representation, and thus the model is not vitiated by Chomsky's observation that children receive little in the way of *explicit* syntactic error feedback and seem resistant to what they do receive. The model requires no negative evidence, and yet successfully hypothesizes a process of dynamic change of an evolving set of word classes and grammatical templates, rather than the all-or-none acquisition of adult grammaticality.

(24)

The learning process is highly dynamic, and what is learned depends upon what has been learned before, so that the same adult input data presented at different times to the model result in different patterns of learning.

Thus, for the learning processes posited here, it is far from true that there is a poverty of data or that the child had no relevant experience. Hearing hundreds of sentences a day and using billions of neurons to do so, surely the wonder is that the child takes so long!

(25)

As support for basing language acquisition on mechanisms incorporating universal grammar, Chomsky argues that to learn whether your language is "head first" or "head last," English or Japanese, it is enough to hear three word sentences like "John saw Bill" or "John Bill saw." If you hear one, "John saw Bill," you have a "head first" language — so it is just a matter of setting parameters in a very highly constrained situation.

From our cognitive viewpoint, this begs a multitude of questions. We must first ask

- "How does the child learn to recognize John and "John," Bill and "Bill," and recognize what action "saw" denotes, and who saw whom?"
- "How is a complex perceptual structure mapped into a simpler structure of words?"
- "How can the child recognize the order of the words in an utterance that its hears?"

(26)

A psychology of language must explain so much of language learning that is not explained by universal grammar that it is moot whether an adequate theory of such learning would leave any lacunae that universal grammar must necessarily fill.

(27)

The model works forward from the evidence that the child provides toward some characterization of the adult language. We ask not only under what conditions a language can be learned by a computational model, but also does the model learn the language in the same way that the child does? To define "in the same way," particular attention must be paid to the "errors" that children make since child speech that differs from adult speech yields clues concerning the processess that the child uses to understand and produce speech. An equally important clue to the child's processes are those errors which the child typically does not make.

One frequently noted phenomenon is the over-generalization of the plural of nouns and of the past tense of verbs in English.

(28)

The Hill Model

The model takes as its input adult sentences together with indications (provided by the modeler, where relevant) of the physical context in which the sentences are uttered. Output from the model is a representation of child-like sentences repeating or responding to the adult input in accordance with the current state of the model's linguisitic capacity.

The child's knowledge is represented by dynamic data structures encoding the child's lexicon, the child's grammar, the conceptual knowledge of the child, and the physical context of the dialogue. The model is given a basic lexicon and a set of concepts with a mapping between the two.

(29)

No assumptions have been made about the ultimate form of the adult grammar nor about what must be built into the model, but a precise account is kept of the knowledge and processes found necessary even for this elementary level of language understanding and production.

Processes attend to the adult input and use rules of salience to focus on examples within the adult data which are used as the basis for language growth.

The input data is in no way especially coded for the model, but is generally taken from language acquisition corpora of adult-child dialogue. The world knowledge is encoded in a semantic net as are the grammar templates and the lexicon.

(30)

The model uses its language experience (i.e., the processing of the input sentences) to build a grammar which is at first a flat template grammar but which eventually evolves into a procedural grammar which may be described, if one chooses, by a set of recursive context-free phrase structure rules. The model embodies 5 assumptions:

1. The child has schemas for and talks about relations.

2. The child has schemas for and employs word order in his utterances.

3. The child employs processes of concatenation and deletion.

4. The child forms classes of concepts and classes of words.

5. The classifying process causes successive reorganizations of the information stored.

(31)

We do *not* assume that the lexical classes of adult grammar are innate.

Rather, we posit a process of *classification through word use* whereby words that are used in similar ways come to be assigned to the same class, thus extending from members of the class to further members of the class certain patterns (templates) of word use.

The initial grammar is given by a set of templates, consisting of a "relation" and a "slot", which is free of any characterization of the adult grammar which will emerge but is not yet present.

(32)

Hill observed a brief stage in which the child concatenated two-word templates with a common word, as in

little bear baby bear

but these soon give way to such three-word templates as

little baby bear.

The four-word utterances with repeated lexical items occurred in such a brief interval that Hill hypothesized that the three-word utterances were arrived at by

• concatenating the two templates "little bear" and "baby bear", and

• collapsing the concatenated relations into a single three-word utterance by deleting the first occurrence of the repeated word.

cf. Matthei 1979 — a young child interprets "the second green ball" as "the ball which is second and green."

(33)

From an adult sentence such as "daddy gave the toy to the boy" the model might initially respond with a single word such as toy. A subsequent presentation of the same sentence might cause the model to acquire a template for gave toy where gave would be classified as a relation-word and toy as a slot-filler. Yet another presentation of the sentence might cause the model to learn the template Daddy gave where Daddy was a slot-filler, and eventually the template (slot1 gave slot2) might be learned for Daddy gave toy.

What is learned in each presentation of the input depends upon the language experience of the model and what has been learned so far. Thus learning is highly dynamic in that each time the same input is presented to the model a different set of grammar rules and additional lexical class information may be learned.

(34)

No information is given the model about word classes, but hearing sentences such as "mommy gave the toy", "John gave the book", "Sue gave the puzzle", would eventually cause the model to put toy, book, and puzzle all together in a word class of words which stand for possible objects of the relation-word gave. Note that it would not matter if the input sentences were far more complex than those used here for illustration. If the model is focussing on the word gave then a sentence such as "Mommy gave the toy to Sue while she went into the store to buy groceries" would have just the same effect as the short sentences used above. The process results in a multiplicity of overlapping and intersecting word classes.

(35)

The model requires schemas for word classification and template classification in order to grow, but the actual classes remain flexible. Processes of generalization eventually also permit the classifying of relation-words which might permit, for example, giving and bringing to be relation-words that could be classed together as words which have similar syntactic properties.

Successive reorganizations of the grammar and the lexicon occur as learning takes place— a process of gradual broadening of word classes and grammatical rules from applying to specific examplars to applying to sets of specific examplars and thence to more general categories.

(36)

Why does the child not end up with an overly generalized grammar or lexicon? It is important to focus on the errors that children make because of the insights which they yield concerning the processes that the child employs in language acquisition. If, however, we permit no overt and specific correction of the child's errors, then how shall we explain why errors of over-generalization do not persist into adult speech?

Consider the verb break. Children at the earliest stage of language acquisition typically learn the word broke but at a subsequent stage of development will start to use the word breaked.

(37)

For a period of time, sometimes for years, both forms exist in the child's vocabulary. How can this period of imbalance between the erroneous and the correct forms be explained? It cannot be explained if the language mechanisms are expressed in terms of explicit rules which the child either does or does not know.

One answer to this question is proferred by our computational model which attaches a weight representing a degree of confidence is with each hypothesis about word forms or grammar rules.

Each time an adult sentence is presented to the model, it is searched for possible instantiations of the available set of hypotheses; and the child's output also represents an instantiation of a particular hypothesis.

(38)

This process of instantiation involves a competition which depends on the current weights associated with the various hypotheses, and that these weights are themselves changed in the process. The weight associated with a hypothesis is increased each time that hypothesis is instantiated in the adult speech input. The weight is similarly increased, but to a lesser degree, each time a hypothesis is employed in the child-like output. In this way more frequently matched constructs are given more weight than less frequently matched constructs.

(39)

Hypotheses must be reinforced to survive. If new hypotheses, however, are to start with very low weights they will have trouble "catching up" with earlier hypotheses. For this reason, separate recency values are employed whose function it is to cause more recent hypotheses to be favored for testing.

A paradigm such as ours may thus be sensitive to the input data and may exhibit varied behavior without the need for negative evidence. Dynamic rule schemas and confidence factors has been used to model the phenomena of generalization, over-generalization, and subsequent correction of over-generalized forms. Thus we need not talk of rules or individual cases which have been learned or have not yet been learned but rather of a continuum in which rule procedures are either strong or weak.

(40)

The model illustrates a manner in which all the input data may be processed, yet with only selected portions of that data focussed upon at different times. By contrast, the theory of universal grammar is based on no attempt to constructively assess the data on, or models of, language acquisition developed to date, and fails to address those issues of language as a medium of communication that are of concern to many cognitive scientists.

To highlight what we have shown, we close by discussing two passages from Hoekstra and Kooij, for they are, alas, typical of the "method" of those who advance universal grammar as a "model" of language acquisition.

(41)

In the first passage, they downplay the argument that a string of words might be judged ungrammatical in English because the pattern to which it conforms is never induced as a possible pattern in English, noting that "this objection implies the assumption that users keep track of the patterns that they have encountered, which is a very strong assumption."

Since this appears to be the basic assumption in our model, we must both defend it and undermine Hoekstra and Kooij's argument.

(42)

• Current versions of universal grammar require that each word have a complex entry in the lexicon, and no claim is made that these entries are innate. Thus the proponent of universal grammar is committed to having the child learn a great number of patterns, for surely it is not too strong an assumption that language users keep track of the words that they have encountered.

• Just as we might not expect the user to remember every occasion on which he had heard a word and yet still to have encoded in memory a "spanning set" sufficient to establish the lexical entry and a set of usages for each word, so is our argument unweakened if we allow the Hill model to have a "spanning set" of patterns rather than a complete set of all patterns ever encountered.

(43)

• The Hill model presents the first stages in the acquisition of patterns of ever greater abstraction and generality, so that the adult's grammaticality judgement is not based on an exhaustive search of every string of words ever encountered, but rather involves rapid access to patterns at the appropriate level of generality through word classes at that level — but word classes built up through experience (our notion of "classification through word use"), rather than given as a priori universals.

(44)

Later on in the same paragraph from which we have just quoted, Hoekstra and Kooij ask us to consider (2a) versus (2b):

(2a) Where did John say that we had to get off the bus?

(2b) Where did John ask whether we had to get off the the bus?

"Both sentences are grammatical. However, while (2a) is ambiguous between a reading in which where has matrix scope or embedded scope, (2b) can only be interpreted as a question concerning the place where John had uttered a particular question. This piece of knowledge is shared by all native speakers, but it can hardly have been established on the basis of induction, simply because there are no data from which induction could conceivably proceed."

(45)

• Where is the evidence that "this piece of knowledge is shared by all native speakers"? It certainly is not shared by all two-year old native speakers. Perhaps Hoekstra and Kooij mean "all *adult* native speakers."

• But even if they are correct, where is the explanation for the transition that occurs in the individual child from ignorance to knowledge? By parameter setting? But where is the explicit description of what the child's language would look like before and after the parameter is set? And where is the corpus collected from a large number of children showing the magical moment at which the crucial datum sets the parameter?

(46)

• We would assert that there is no one crucial datum. Rather, the child is exposed to a vast array of data which include questions whose patterns are reflected by (2a) and (2b). The very patterns that enable the child to master the words "ask" and "say," "that" and "whether" are the same patterns which give the child "the data from which induction could .. proceed" whose existence Hoekstra and Kooij deny.

They deny it without *any* empirical analysis of how language changes with experience. We have *shown* that simple patterns can evolve into complex patterns in a way which matches patterns of language acquisition in a two-year-old child.

(47)

We *claim* , as a target for future research, that our model can be extended (not by the formation of word patterns alone, but through the continuing interaction of the lexical space, grammar space, and cognitive space of the child) to cover such phenomena as the distinction between (2a) and (2b). An adult (trained in linguistic terminology) can learnedly discern patterns in (2a) that can be distinguished as matrix scope vs. embedded scope, but this adds no weight at all to the claim that such knowledge is embedded in an innate universal grammar, and that without "knowing" matrix scope vs. embedded scope innately, the child could not acquire the ability to distinguish (2a) from (2b). In our society, children initially learn language through using it without any necessary reflection upon its patterns.

(48)

Eventually, the child does come to reflect with pleasure on these patterns, as well as to experience, with less pleasure, the explicit presentation of grammatical rules in the classroom. But this does not concede an explanatory role for universal grammar in language acquisition. We argue that the rules are structures whose acquisition is made possible by the prior acquisition of language, not innate structures that make the acquisition of language possible.