Brains, Robots, and the Evolution of Language[1]

Michael A. Arbib
Department of Computer and Information Science
Technical Report 74C-2
University of Massachusetts
Amherst, Mass. 01002

ABSTRACT

This paper presents an outline for two courses:
"Cybernetics and Language", to be offered at the Summer Institute of the American Society for Linguistics, University of Massachusetts at Amherst, Summer, 1974.
"Computational Cybernetics", to be offered as COINS 782 at the Department of Computer and Information Science, University of Massachusetts at Amherst, Spring, 1975.

Incisive criticism and suggestions for material to be added for either course will be most gratefully received.

---

[1] This paper is based on lectures given to the Sigma Xi of Mount Holyoke College on May 8, 1973, and to California State University at San Jose on May 14, 1973. Preparation of this paper is supported in part by NIH grant number 5 R01 NS09755-04 COM.

1. INTRODUCTION: Some Points of Intersection for Cybernetics and Linguistics.

("How can we talk to a Robot?": A question about computer science and a question about the human brain.)

Let me briefly present reasons why someone who wants to understand information processing in complex systems might want to know about linguistics.

Linguists, in trying to understand languages, have come up with various types of formal description. Chomsky's formal descriptions of grammar have proved helpful in letting us talk about programming languages, the formal language systems that we use for communicating with our machines. However, here I wish to emphasize two other themes. Firstly, one of the big controversies in linguistic theory at the moment is the relation between syntax and semantics. Syntax involves such problems of formal grammar as "Is this sentence grammatically correct?" as in "Furious ideas sleep greenly", which may not mean very much, but is grammatically correct. Semantics emphasizes meaning. I think that until now many of the key questions of semantics have been badly neglected by the linguist, and that we begin to get a handle on it in the attempts we are now making to communicate with computers using approximations to natural language. We shall take up this topic in Section 3.

We shall study two themes relating the attempt to understand the brain to attempts to get a better handle on language. Firstly, looking at some of the commonalities between how we perceive and how we might expect a robot to perceive will suggest how language can be seen as evolving within this sort of perceptual basis. Secondly, we shall discuss the effects of brain damage upon the ways in which people speak, and suggest that this gives us a view of the brain as a highly parallel computer.

2. PERCEPTUAL FUNCTIONS COMMON TO ROBOTS AND ANIMALS, AND HOW THESE FUNCTIONS PROVIDE A NATURAL BASIS FOR THE EVOLUTION OF LANGUAGE IN HUMANS.

2.a What Does an Action-Oriented Perceiving System Need?

In this section we discuss what animals and robots need as they interact with the world. The first requirement is a spatial framework, an idea of where things are in the world. The second need is some sort of segmentation, the ability to break the world up into different "chunks", as when one enters a strange room, and recognize that there's a chair there, a table here, a friendly robot there, and so on. And thirdly (but this list is by no means exhaustive), one has to have long-term and short-term models of the world.

2.a.i. The Spatial Framework

(Frog Tectum; Bower on Infant Visuomotor Behavior)

A perceiving system is, for me, one which must interact with its world. It thus needs a spatial framework--if I want to put my keys down, and recognize a surface on which I may place them, this recognition is invalidated if by miscomputation I have the metric relationships wrong and so drop my keys on the floor. Fortunately, the way in which our brains are genetically "wired" provides a basis for spatial interactions. For example, the tectum (a region in the midbrain) of the frog receives a map of visual input--if we measure electrical activity at different points of the tectum in response to visual stimulation, we can use this to map out the visual field. If we electrically stimulate a small region of the tectum, the frog snaps at some point in space, and this point is in precisely the direction for which visual stimulation is optimal for that region. In other words, here we have a spatial layout in the brain which gives the animal the relationship between its input spatial framework and its output spatial framework (where it must zap to get the fly).

If such a spatial relationship occurs in a system as relatively primitive as the frog, one might expect to find it wired into the brains of human infants. In fact, Bower, Broughton and Moore [1970 a,b] found that even in the first week of life, a baby could reach for a seen object (although inaccurately) and could focus its gaze upon its hand when no object is present. Further, the baby does not have to alternately fixate hand and object to guide the hand to the object, for the baby can swipe as accurately if lateral blinders prevent him from seeing the initial position of the hand until a ballistic movement is initiated. Thus, we now have evidence that a similar mechanism to the visuomotor coordinating mechanisms of the frog is apparently genetically specified in humans. (Trevarthen [personal communication] notes that eye-hand coordination is, however, initially slow and relatively imprecise; but that there is a rapid and regulated development which apparently involves the better integration of controlling mechanisms by feedback from practice.)

To add to this evidence for an inbuilt spatial frame, we should mention evidence that we have evolved to perceive objects rather than sense intensity of stimulation or size of retinal stimulation as isolated variables. Bower et al [1970 a] show that infants only 1 to 3 weeks old will already show more of an avoiding response to an approaching object than to a puff of air alone, or to a movie projection of the approaching object; while earlier Bower had shown that a 2 month old infant will respond to a cube at different distances as being more alike than a small cube nearby and a large cube far away which yield the same retinal image. In robotics and traditional psychology we often seek the minimal set of cues sufficient to identify an object. However, the above data suggest, but do not prove, that the brain makes judgments on many variables rather than on one variable at a time.

In another experiment, Bower found that 2 month old infants were distressed if a different object emerged from one they had just seen disappear behind a screen, suggesting that some aspects of object constancy appear early. This would seem to me to fit in well with the notion, to be elaborated in Section 2.a.iii below, that we perceive an object by setting up an internal model which is easily modifiable to adjust to the movements of the object. If basic transformations of simple objects are as basic as we might expect them to be, the baby's surprise would not be surprising. Presumbably, then, one aspect of mental development is the ability to model objects of increasing complexity, and "hold on to" these models for increasing lengths of time.

It does seem, then, that biological systems are provided with some sort of innate spatial framework within which their perceptions may grow. By the same token, when we build a robot we build into it devices for giving it very accurate spatial control; we give it a television camera equipped with a range finder so that it may identify just where objects are in space; and we provide it with counters whereby it may keep track of how many steps its wheels have turned so that it may correlate where it has to go with how far it will have to move to get there.

2.a.ii. Segmentation and Analysis-by-Synthesis

(Dev on Segmentation; Boylls on Synergies; Hapsis.
Hierarchies and Complexity: Simon on Evolution; Structured Programming.)

The next perceptual ability we all have from birth and which we have to put into robots is the ability to segment a complex scene into its pieces. When one comes into a new environment, to analyze it ab initio one must first separate the figures from the background as part of the process of recognizing the individual people and objects, etc., which constitute the environment.

We might say that we analyze our world by synthesizing it in our heads. Suppose, for example, that I find myself in an auditorium with people in it, without knowing who the people are, where they are, etc. Slowly I segment out different parts of my visual input and manage to make sense of these parts (getting to know where different people are) to build up in my brain a pattern of activity that "takes care of" more and more details of the world around me. This is not to be thought of as forming a "kodachrome" in my head: the interesting thing is not that there is some pattern of brain activity that fully reflects the patches of color that constitute the surfaces "out there"; but rather that I have formed an internal representation, built out of perturbations of the internal representations of familiar things, to make sense of the input. I have in some sense analysed my environment when I have synthesized a useful representation of it in terms of what I already know.

One of the things we have been doing at the University of Massachusetts in our Cybernetics group is looking at a number of brain models related to segmentation. Parvati Dev has been looking at how a mechanism using components that look like the neurons of the visual cortex could take a scene which is initially just, to use William James's phrase, "a buzzing, blooming confusion" of neural activity, and slowly segment it out into different regions, without any appreciation of what the regions mean yet, but simply in terms of "this is foreground", "this is background", etc.

On the output side, this idea of breaking things down into pieces which are relatively familiar has been extended by Curt Boylls in his work on the cerebellum, that "little brain" attached behind the brainstem which seems to play a role in the coordination of movement. The paradigm for his study is the work of the Russian school founded by Bernstein: rather than describe how a horse is moving by giving, for each of several hundred muscles, the state of contraction of each of these muscles at various times, one can say, "The horse is cantering" or "the horse is galloping" or "the horse is trotting". The suggestion of the Russian school, substantiated by their experiments, is that in some sense the neural control of movement proceeds far more in the second mode than in the first. In other words, the brain does not try to independently control the tension of all the muscles in the body, but rather instead has a number of gross strategies--which we call synergies--available to it. It gets the body set into the right "ballpark" of movement and then tunes the parameters within those "ballparks".

Work on hapsis brings together the input and output aspects of segmentation. In building up a knowledge of something by touch, one feels its contours, and slowly over time gets to know where the corners are and how they're related. In this case, one is actually using one's familiar output routines such as running one's hand along something or feeling around a corner to provide the building blocks for an internal representation which provides the analysis of what one feels as a synthesis of the movements which allow one to economically explore it.

This idea of not trying to handle the buzzing, blooming confusion all in one go but by breaking it down into familiar pieces is illustrated in so many other areas that there is almost an embarrassment of riches of things to say here. In computer programming we now have the idea of structured programming--the fact (so obvious it's a truism) that if one wants to write a program which has 30,000 instructions one should not try to write down 30,000 instructions one after another and hope the program will work; but one should rather break the overall problem down into meaningful sub-problems and keep breaking those down, with maybe a little bit of "tuning" at the end as you relate the internal structure of some of the blocks. It is only with this structure of large blocks in interaction that we will have any chance of understanding the computation.

Again, if one looks at evolution, it would seem that animals have evolved in terms of some fairly "grand" overall strategies--fish vs. bird vs. mammal vs. reptile--within which further differentiation to match specific conditions takes place, rather than a completely uncorrelated exploration of every ecological niche. I think the point is made that whenever one studies complexity, and certainly in perception--be it of robots or of humans--the process of segmentation is going to play a crucial role.

2.a.iii. Models of the World

(Long-term and Short-Term; Context; The Slide-Box Metaphor; Notes on Adaptation; Realization Procedures.)

We have already noted that it is not enough to be able to recognize a surface which should support the weight of some object, but one must also have a spatial framework within which one could position that object. Unless one tries to program a computer to do it,it is hard to be impressed by the fact that when one comes into a strange room a pattern of highlights different from any previously encountered can, with "incredible" predictive power, be "instantly" perceived as a surface which will support the weight of the keys. I would characterize this ability as the possession of a long-term "model of the world". As we grow up in the world, we come to be able to recognize objects in terms of rather minimal cues. One dramatic example is that a person who has grown up on the veldt of Africa will eventually look at the long grass in the distance and see what you or I would take to be a slight shift in the wind direction and say, "There's a lion". Note that they do not consciously think "I see that particular type of bending that is not correlated with wind gusts, but is in fact, as I know from experience, correlated with lions"--instead, they see a lion. The difficulty for us is to get far enough back to see that it is as amazing to say "I see a lion" when we see the shape of a lion without having such dramatic forms of confirmation of our hypothesis as to actually feel its claws ripping into our flesh.

By contrast, let me mention the short-term "model of the world". By this I do not mean some sort of reverberating message which sticks around in electrical form for 15 seconds before being transmuted into DNA and RNA. Rather, I mean that at any time, in addition to what one is in some sense receiving in the way of immediate stimuli, there is a lot in one's head that tells you what is "going on". When facing an audience in a classroom, I know that if I turn around there will be a blackboard, even though I receive no stimuli to reinforce that part of my awareness. To try and get some feel for this phenomenon, I've coined what I call the "slide-box" metaphor. In drawing movie cartoons, a common technique is, rather than to complete a new drawing for each frame, to specify a few parameters--lower the arm so many degrees, shift that tree so many inches--to indicate how to reposition the "slides" which constitute a frame to obtain the next frame. Moreover, in these cheap cartoons one may use many pictures again and again--one might have a file of slides containing, among other things, six different perspectives on the family car which suffice for nearly all scenes in which the car appears.

To provide a metaphor for our own short-term modelling of the world, consider how one might make a film of an actual event using these slide-box techniques. We do not use a conventional camera which represents the light intensity of every point of the scene, but knows nothing about what objects are in the foreground, which are in the background, or what they are. Rather, the idea in making a cartoon representation is to find out whether and where there is a car in the picture, where the heroine is, whether she's asleep or awake, and a few other exciting and important details. The job of the camera is then no longer to set up a complete representation of the light and shade, but rather to help retrieve the appropriate slides from the "slide-file". [Let's see, there's a wheel there, now where's my box of things with wheels on, ok, here's a car, no, that doesn't quite fit, ... rummage ... here's another, that's X's car, rather than some other car] and so on down the line until eventually one has retrieved slides which can be put together to let us say "That's close enough, now I know what the important things are out there". Once you've done that hard work of the initial recognition of a scene, from there on it goes very smoothly; one usually has to make fairly small changes unless one has goofed dramatically, or unless somebody new comes into the scene.

That we do perceive by "modelling" is illustrated by the fact that when one is walking down the street towards someone far away, one may first go through the blur period, then through the "I'm sure it's Y " period, and then the person comes close enough for you to see further details that destroy that perception--even though for a while you did not think "Those features seem to be similar to Y's" but "I see Y vividly". In other words, at that stage, (and I hope you'll excuse the crudeness of this metaphor) you have activated in some fashion within your brain, a pattern which in the past has been correlated with Y, and it has met the current input so satisfactorily that you perceive the approaching person to be Y.

2.b. Evolution of Language

In discussing Noam Chomsky's views on the development of human linguistic ability, Lakoff [1973] points out that:

> Chomsky has claimed that people possess innately not merely general learning mechanisms, but a specifically *linguistic* innate faculty. His argument is of this form: There are complex linguistic universals that everyone learns uniformly. There are at present no general learning theories that can account for this. It is hard to imagine what any such theories could be like. Therefore, it is plausible to assume that there can be no such theories. But the argument is fallacious: Nothing follows from a lack of imagination.
>
> What Chomsky has shown is that either there is a specifically linguistic innate faculty *or* there is a general learning thoery (not yet formulated) from which the acquisition of linguistic universals follows. The former may well turn out to be true, but in my opinion the latter would be a much more interesting conclusion. If I were a psychologist, I would be much more interested in seeing if there were connections between linguistic mechanisms and other cognitive mechanisms, than in simply making the assumption with the least possible interest, namely, that there are none.

My purpose in this section will be to take up Lakoff's challenge by arguing a plausible scheme for how language might have arisen as an elaboration of more fundamental functions of perception. This does not deny that the evolution of language has greatly enriched human behavior, but does deny its primacy. (For a more philosophical discussion of the secondary role of language, see Arbib [1972].)

In Section 2.i, we have studied some of the basic perceptual abilities required for an organism which interacts with its world: a spatial framework, within which to locate things in relation to the ability to act; segmentation, the ability to break the world into "pieces"; and the "models" which give meaning to those "pieces" in terms of previous experience, involving both the long-term accumulation of knowledge of what different types of stimuli can mean in the possible ranges of interaction that are afforded by the environment when those stimuli impinge upon the organism; and the short-term, given metaphor as the current contents of the "slide-box", the current internal representation of the environment. [Long-term models in the case of the robot's computer are programs, which can take features sensed by the TV camera to start pattern recognition and object recognition routines. The short-term model is the representation within the computer of all the objects in its environment that it has currently been able to recognize, these stored in such a way that the computer is then able to act.]

Now, although I've had to use language to communicate them, and although the program of a robot would be expressed within its computer in some formal programming language, these perceptual abilities are all prior to language. The ability to see what objects are in the world, to discriminate one from another and to be able, on this basis, to interact appropriately with the world is shared by all animals, though animals differ markedly in what they can discriminate and how sophisticatedly they can interact. A frog will starve in the presence of dead flies and repeatedly zap at a friend scratching his back, the moving digit providing a similar stimulus to the wiggle of a fly. A cat is already far more sophisticated in what it can discriminate, and humans are more sophisticated yet. I want to suggest that language is not a magic, separate device to be explained by formal grammars, but is rather an ability which evolved naturally out of our ability to perceive the world.

2.b.i Call Systems

(Bees and Birdsong; Bennett on Rationality.)

A given situation or object can be re-cognized (known again) in many different ways. But given this ability, which humans would seem to share at least with all mammals, the ability to use signs seems almost immediate. Many organisms, knowing an animal by its appearance may come to know it by its cry; and in those cases where the brain is so structured as to enable the organism to make this association on the basis of experience, then the cry might as well be an arbitrary grunt or howl. Thus the ability to recognize signs would seem to be on a par with the ability to perceive the world in terms of objects (yielding "noun" recognition) and interactions with them (yielding "verb" recognition).

What distinguishes humans, it would seem, is then not so much the ability to recognize signs as it is the ability to assign them to new simplexes and combine them to generate complex sentences. [Studies of the speech of brain-damaged humans help us probe the mechanisms involved, as we shall suggest below in citing studies by Geschwind [1970] and Luria et al [1970] on language in relation to cerebral organization.] In trying to suggest what stages might be intermediate between the primitive ability to perceive and the sophisticated use of language, one might state with very primitive call systems, such as the babe's crying (a very specific signal to attract the attention of the mother) or some of the noises accompanying courtship rituals. Such calls are stereotyped and rather small in number; they fit rather gross situations. Birdsong, again, usually has some rather limited expression in most species, and is tied with rather specific types of territorial activity. The interesting move from a few rather gross signals, a rather crude call system, to greater differentiation can occur in at least two different ways: via continuous gradations and via blending.

A closed call system has one call for each discrete type-situation; whereas an open system is one which, like that provided by a generative grammar, has a potentially infinite "call" set. Two ways to open up (introduce fine gradations) into a call system are given by

a) Associating continuously variable features of a single call--such as pitch, volume, or duration--with continuously variable features of a type-situation--such as degree of danger, quantity of food, direction and distance. This is logically like the bee dances, but unlike human languages.

b) Blending: Combining 2 calls to signal complex situations. This mechanism starts opening up the call system by providing the basis for the habit of building composite signals out of meaningful parts, whether or not those parts occur alone as whole signals. [I again stress that the building of composites has a substrate in prehominoid building of internal models].

One might imagine that the intensity or pitch of a call might in some sense signal the intensity or pitch of emotion it is to convey. Perhaps the most dramatic graded system of what we might call pre-language is in the dance of the bees. When bees return from a successful discovery of pollen, they can carry out an intricate dance within the hive, and the angle of this dance and the rate of wiggling can signal not only that food has been found, but in what direction from the hive it is and how far away it is, and even something about the quality of the food. This is already well beyond a grunt or two, and yet it lacks the flexibility of language. Although the bees can be fairly precise in their specification of these parameters, they can't get much farther than that. This observation provides the basis for a little book by Jonathan Bennett entitled "Rationality" in which he posits pseudo-bees with greater and greater powers until he is finally prepared to concede that they're rational. They must have the ability to lie--that seems very important in language, otherwise where would science be?! If Einstein hadn't had a linguistic system in which he could say things which the majority of scientists in his day "knew" were patently false, we would not have the theory of relativity today. The ability to state hypotheses which may or may not turn out to be true, the ability to deny things, the ability to call for evidence and evaluate it are all important aspects of our linguistic ability which distinguish it very strongly from the signalling system of real bees. Speculating in this vein, we might suggest that, initially humans used conventional cries to draw attention to a single object, or to encourage the listener to undertake some particular action. It would then be natural to suggest an action with respect to a particular object and so concatenation--the chaining together of signals--could then evolve. Concatenating the typical warning cry of the mother to her child with the signal for some forbidden action might then be the basis for negation, and the first step toward logic.

Let me close this brief discussion with two observations on comparative studies. Bronowski and Bellugi [1970] analysed the attempts of Gardner and Gardner [1969] to teach chimps sign language. Earlier attempts to teach chimps to speak had failed, but the Gardners had realised that this might reflect an inadequate control of the vocal musculature in the chimpanzee rather than a lack of linguistic ability per se. In fact, using sign-language the chimpanzee was able to reach what Bellugi and Brown [1964] characterized as Class 2 in the linguistic development of children. This shows that while those aspects of language peculiar to speech are not shared by the chimpanzee, such basic linguistic abilities as the use of concatenation are. Complementing such studies, Marler [1970] asks "Birdsong and Speech Development: Could there be parallels?", for while birds do not have concatenation or language they can learn complex sound patterns. Marler shows ways in the analysis of the relative degrees of genetic fixedness and learnt "dialect" in various types of birdsong may provide valuable cues in probing the evolution of, and brain mechanisms for, our own speech abilities.

2.b.ii. From Prelanguage to Language

(Blending and Open Systems; Romer's rule: Trees, Carrying, Manipulation, Tools, Free Mouth.)

A crucial possession in passing from the bees to the scientist, is the open system in which previous calls can be put together in new combinations, and the compound calls can in turn come to have new meanings. For example, in English, a pen is something for writing with, a knife is something for cutting with; put them together and you've got "penknife" a knife ideally suited not for writing but for sharpening quills, a type of knife which still has meaning in this age of ballpoint pens. This ability to put things together, and then for this new combination to take on meaning of its own seems to be a crucial development in the history of the evolution from a simple calling system to a flexible language.

What I wish to emphasize is that this crucial step is clearly foreshadowed in the development of the animal's perception. We've stressed that the ability to break the world down into relatively recognizable pieces is something which we share with far lower animals. The ability to put those pieces together into a short-term model and thus make use of <u>context</u> is crucial. A cat realizes that a certain piece of fish in one context is food, but in the context of a dog standing over it is something to be run away from. Yet if the cat were very hungry, or the dog were very small, then the fish would be food. So the ability to put things that are known in isolation into combinations and give different meanings to the parts dependent upon the partners is not unique to language, but is already preshadowed in our perceptual systems. Thus, the evolution of language does not require completely new mechanisms but rather is a somewhat more abstract recapitulation of things that we've already seen in the evolution of our perceptual abilities.

To spell out these ideas, we give an account of human language evolution, based on Hockett and Ascher [1964], but amended to stress the commonality, missing from their account, between language and perception. Their essay attempts to set forth the story of the emergence of the first humans from their prehuman ancestors, but we shall only cover that portion concerned with the evolution of language. They start by considering the descent of hominoids from the trees, noting that

> The conditions for carrying are no better on the ground than in the trees if the hand must revert to the status of foot. But if bipedal locomotion is at all possible, then the hand is freed for carrying; and the survival value of carrying things [pre-tools; scavenged food] in turn serves to promote a physical structure adapted to bipedal locomotion.

They repeatedly invoke what they call Romer's rule: "The initial survival value of a favourable innovation is conservative, in that it renders possible the maintenance of a traditional way of life in the face of changed circumstances." They then note that

> ...carrying made for a kind of behavior that had all the outward earmarks of what we call 'memory' and 'foresight': one lugs around a heavy stick of stone despite the absence of any immediate need for it, as though one were remembering past experiences in which having it available was important, and were planning for possible future encounters of the same kind. ... the outward earmarks [of 'memory' and 'foresight'] surely came first, and only over a long period of time produced the psychological characteristics to which these terms refer.

However, note that we view memory or planning in terms which apply to prehominoids. The point is better made, perhaps, if one views the authors as using these terms at a level closer to that admitting conscious verbalization in humans.

> The use of the hands for carrying implied that the mouth and teeth, classically used for this by land mammals...were freed for other activities. ... Remember that the proto-hominoids are assumed... to have had a call system. ... The hunting of dangerous animals is a challenge ... [and] there is a great advantage if it can be done collaboratively. But this calls for coordination of the acts of the participants. Their hands hold weapons and are thus unavailable for any complicated semaphor. Their visual attention must be divided between the motions of the quarry and those of the other participants. All this favours an increase in flexibility of vocal-auditory communication... Collective hunting, general food-sharing, and the carrying of an increasing variety of things all press toward a more complex social organization, which is only possible with more flexible communication.

Hockett and Ascher then proceed with a view of language evolution based on blending:

> The young may emit some of the calls instinctively. But they are also exposed to various more or less complex calls from their elders, and are obliged to infer the meanings of the parts, and the patterns by which the parts are put together to form the whole signals, from the acoustic resemblances among the calls they hear and from the behavioral contexts in which they are uttered. Thus, the development ... [of blending] puts a premium on any capacity for learning and teaching.
>
> If the conventions of a system have largely to be learned before the system can be efficiently used, then much of that learning will eventually be carried on away from the contexts in which the utterances being practiced would be immediately relevant... [adding] displacement ot the [other] design features. ...

We have now reached <u>prelanguage</u>: an open system, with details transmitted largely by tradition... and with the property of displacement, but lacking <u>duality</u>: the step from treating calls as wholes, via paying attention to details to distinguish different calls, to interpreting these acoustic details as identifying or representing morphemes. The development of a refined articulatory apparatus came about, in man, through the marked separation of glottis from velum--but note that this discussion of language is at a more fundamental level than would require speech production to be restricted to a specific apparatus [which ties in well with the point of recent studies on teaching chimpanzees to "speak" (language or prelanguage?)] These developments "rendered possible the continued use of a thoroughly familiar type of communicative system in a thoroughly familiar way, in the face of a gradual but potentially embarrassing increase in the complexity [i.e., increased vocabulary] of the system"--Romer's rule, again.

2.b.iii Generative Properties

(Chomsky vs. Skinner: Innate Ideas vs. Stimulus Response; Representation and Reality: Zeno and Suppes. Communicating about "Slides": Sentences and Ballparks. Moving from semantics to pseudosyntax to reduce ambiguity; Semantic Version of Thorne's parsing scheme.)

Just about every sentence we utter is one we have never uttered before. Many modern linguists, notably Chomsky, have stressed that this implies that language could not be learnt in a naive stimulus-response way which involves rote learning for each individual string of words, whether it is grammatcially correct or not. However, this does not imply that the learning of abstract grammars is a prerequisite of language ability, for it ignores the more abstract level suggested to us by the slide-box metaphor of 2.a.iii. We have learnt how to analyze a scene into its parts and put those parts together meaningfully. If we think of language not as some string of words written down in a book for a linguist to analyze and detect regularities in, but rather as evolving as a tool for more and more effective communication, we ask: what is being communicated? Presumably some aspects of the internal model of the world which we wish to share with someone. There you are back in the old days, running around a mastodon, jabbing your spear at it when you see that it's about to veer away, and you want to yell to Harry Og on the other side to do something about it. Now as your ability to communicate some of your perceptions about what to do with this mastodon increases, then it becomes possible to develop new refinements in your symbol system. So we get this sort of hand-in-hand evolution of ability to communicate more and need to communicate more beyond that. Thus this generative propery of language which has seemed so mystical to many grammarians seems a very natural consequence of our thinking about perception in terms of breaking things down into familiar pieces.

Let me illustrate this secondary role of syntax in another way: if I say,

"apple", "eat", "boy"

then without any syntax there it is pretty clear that what I am talking about is a boy eating an apple. But strings like

"girl", "chase", "boy"

are hard to disambiguate. Here it's not enough to have the straight semantic components, but one must also indicate their relationships. Then again, what one is trying to do is express what's actually going on, what is in the "model" that you would make. Going back to our slide-box metaphor which is a bit too visual, but will do for our purposes, we have quite a different picture if the girl is chasing the boy, or if the boy is chasing the girl. In the same way, one expects that our brain's internal representation captures such a distinction. In translating this internal "model" into a word-string, one has to have something in the strings of words which takes the same components and yet expresses the different relationships between them. It seems to me better to think of this in terms of helping clarify semantic relationships, meaningful relationships, than it is to think of the resultant rules as constituting a grammar which plays a primary role in generating structures to which meaning is attached at the last moment. The child discovers regularities in language, and then uses these to <u>resynthesize</u> a sentence. For example, a child of 2 may understand "the cat was chased by the dog" as "the cat chased the dog". I would interpret this as showing that a child does not parrot the sentence but rather understands it (builds an internal model of what he thinks it says) and then expresses the model in a sentence with a structure he already knows.

To get a feel for the sentences the child produces, we may study how the child develops the use of negation:

I. At first he puts no or not in front of a sentence to be negated.

After a time he utters such sentences as

II. I no do it.
He no want it.
Mummy no cut it.

The negative element is placed inside the sentence--not as a direct imitation of adult sentences but rather as a reconstruction according to the child's own rule. When asked to repeat a sentence such as "I don't like it", a child of this stage (around 35 months in one case) pauses a while before saying "I no like it".

Later, (3 months later for the child mentioned above) the negative becomes associated with the auxiliary system.

III. He don't want it.
I can't have it.
I don't want some.

Such grammatical refinements (fine-tuning) as changing "I don't want some" to "I don't want any" do not come till later. It should also be noted that including a negative can "take up room", and so may yield "economised forms" as when

"Catherine have shoe on" is negated to yield "Catherine no shoe". Thus, besides acquiring new rules, the child increases "capacity". A child may have more to say than he can express. Between stages II and II (the stages may be 2 or three months apart) the child seems to use "It's" as a variant of the pronoun, not as a combination of 2 elements, as in "It's goes". I.e., the child may segment its sentences differently from the way adults do.

Why does the child change his grammar toward that of the adult when he can already make himself understood? Cutting down the number of rules accounts for some changes, but not for "fine tuning" like "who" versus "whom". Perhaps after the child has learnt how to express an idea so that he can be understood, he is then free to learn the "fine tuning" of increased grammaticality which will increase his circle of listeners by imitation. In any case, we repeat the observation made above that the child does not parrot sentences but reconstructs them according to his own rules. This reinforces our objections in Section 2.a.iii to the seductive simplicity of a "Kodachrome" model of human memory, which regards memory of an event as some sort of faithful encoding awaiting retrieval at some later time. I would suggest that it will prove more fruitful to regard the fabrication of a unique and absolute representation of each experience as neither economical in neuronal terms nor flexible in dealing with novel stimuli, and that the crucial task in learning may be seen as the discovery of which features or rules make different events usefully similar.

3. PROGRAMMING COMPUTERS TO "UNDERSTAND" NATURAL LANGUAGE

3.a. Representing Sentences as Procedures

(Winograd; Action-Oriented Statements.)

The most interesting feature of Winograd's approach to getting computers to "understand" natural language is that it makes essential use of the notion of a model of the world, so that sentences are analyzed not so much by grammatical ideas of syntax like locating nouns or adverbial phrases, but rather by translating a string of words into a program which the computer can follow to obey commands. In Winograd's thesis, a 200 word vocabulary requires 80 K of storage. Thus a necessary step was to find a "world"--in this case a block world--complex enough to allow the robot[†] to do interesting

---

[†] Winograd did not actually build a robot and have it interact with a world with blocks in it, but rather he simulated a robot interacting with a simulated world. For vividness we will talk as if there were a real robot in a real world, for we're interested in the way in which the "robot" can understand language to do our bidding.

things, and which would nonetheless be tractable with reasonable computing power. The "world" consists of a number of blocks, some piled on top of each other, of different shapes, sizes and colors. The robot, in trying to analyze a sentence like, "Pick up the red block and put in on the green block" is, essentially, trying to turn it into a program to control its actions. That means that it must locate where the red block is--which will normally involve reference to an internal model built up during earlier analysis, rather than an actual searching around the world. If, in this particular case of "Pick up the red block and put it on the green block", the robot finds it has two red blocks in its (model of the) world, it cannot generate a unique program to guide its activity and it will then call for the requisite information by asking "Which red block?" to which you might reply "The one on the left" which would disambiguate your command; at which stage the robot could put together the appropriate program to control its response.

This language behavior is clearly not based on an abstract analysis of the sentence, but is rather based on the use of the sentence for controlling its behavior given its current world. The response "which block?" would not have occurred if there had been only one red block in that particular environment. And so it goes on. We have here the beginning of a framework in which we can have a feel for how to think of language in a natural way in which the communication of action-controlling or action-relating functions is made primary, and where syntax enters purely in a secondary role.

3.b. Grammatical Inference

(Harris, Chomsky's Innate Ideas, Realization Theory.)

Harris's approach to computer understanding of natural language shares with Winograd's the idea of having the machine talk about a model. However, more effort is devoted to enabling the machine to build up its language over time. In the initial phase, the machine starts with some internal concepts and, by simple associations, learns the English words for them. More interesting is the second phase, which gets the machine to build up a grammar over time. The approach here is very much in the style of Chomsky's innate ideas, because Harris gives the machine the "idea" of a class of grammars and some strategies to build up grammars which match the data. The system would not initially know anything about sentences, noun phrases or verb phrases but does have built into it the idea that a grammar is used by starting with a symbol and making successive replacements using a family of "productions", or replacement rules, with symbols so obtained being replaced in turn by other intermediate symbols until finally obtaining a string of terminal symbols which constitute a grammatical sentence. It then has a number of procedures for generating grammars which match the given sentences, one of which is to lump together a couple of symbols to form a new intermediate symbol. It wouldn't use the symbol "noun phrase", but might designate some new symbol A as something which can be replaced by the pair of symbols 'the dog' that it already knows. Similarly, another symbol B might be replaceable by 'ate tarantulas'. As time goes by, it would have a number of symbols which could be replaced by other symbols. It might eventually discover that it could get a more economical description if A could also be replaced by 'the cat'. In this way, it would build up, through a number of operations I shan't go into, a relatively small number of symbols via which it could derive all the sentences it had been exposed to, and could further take account of a lot of new sentences that came in as well.

Interesting though this is, I would not wish it to be thought of as supporting Chomsky's idea that somehow a newborn human has embedded in its head a magical formal-syntax-generating system that has simply to be tuned by experience. I have tried to suggest a more semantic procedure that starts from the ability to perceive the world in certain ways. A child starts off with one-word 'sentences' like "Milk", and it's up to the parent to determine whether that means "I want a drink of milk" or "Take the milk away". At the next stage, the child can master two words at a time, and some linguists have managed to invent syntactic categories and write grammars for the two year old. I think it's much simpler to say that when you get to a certain stage you can manage two things at a time, and one's going to be a bit like a noun, and one's going to be a bit like a verb, and not make anything big out of it. This building up of more and more complicated sentences is very different from Harris's scheme, which starts with full-fledged sentences and attempts to build a full-scale syntax for those. While stressing its differences from the way in which I believe a child acquires its ability to communicate, I do think Harris's program is an important development in our understanding of how we can build up economical descriptions of our world by "folding together" and "tuning" existing structures--and bears careful comparison with Piaget's "accomodation" and "adaptation".

The point of these comments on some of the attempts to get computers to understand English is to stress their primary use of the existence of a world about which the system is communicating; and to suggest that while one can use syntactic structures for these systems, they seem to be of secondary importance to the internal-model-communication aspects.

4. Some Effects of Brain Damage on Verbal Behavior which Suggest a Model of the Brain as a Highly Parallel Computer.

4.a. Neuroheuristic Programming.

Many of the most interesting programs in artificial intelligence are based on heuristic procedures which need not be guaranteed to work, but which do incorporate a lot of good ideas which seem to yield interesting solutions in many cases.

In thinking about brains, I'm very interested by what we might call neuroheuristic programming--the idea that one doesn't simply look for a program that solves problems in an interestingly heuristic way, but does it in such a way that the various pieces of the program can be correlated with various pieces of the brain.

4.b. Gross Damage to the Brain.

Years ago, I read a Pogo strip in which the bear had been writing some poetry. Pogo came up and asked him what he had written. The bear answered that he didn't know because he couldn't read. I though that being able to write but not read must be a fantasy, but a most intriguing result on the effect of brain damage upon verbal behavior was found around 1890 in Germany which revealed just that effect!

As you know, the cerebral cortex is divided into two hemispheres, and these hemispheres communicate both in terms of lower structures of the brain, evolutionarily older structures, but also in terms of a massive pathway called the corpus callosum. In the majority of people, most of their "linguistic mechanisms" are in the left hemisphere. The subject of the 1890 investigation had massive tissue destruction in the visual cortex on the linguistic side and that part of the corpus callosum that kept the two visual cortices in communication. To speak very crudely, the subject's language centers still had the forward portion of the corpus callosum whereby to communicate with motor activity on the right half of the brain, and also of course could communicate directly with motor activity on the left half of the brain. So he had no trouble with directing his hands to write--moreover, at the level of shaping letters the other hemisphere was able to supply appropriate visual feedback. But when it came to reading something, only visual input to the left half of the brain would suffice, but the visual cortex on the left side was destroyed, and there was no way of getting information across from the right hand side. Thus, the patient could write but not read.

4.c. Access to Subroutines

(Luria)

A. R. Luria studied defects in writing caused by lesions of the brain. He found one brain site concerned with 'phonemic' organization: b and p 'sound' very similar and, with a lesion in this site, a patient will make mistakes in writing such as interchanging b and p. A second site is concerned with 'kinesthetic' organization: l and d 'feel alike' in the mouth of the speaker and with a lesion in this site a writer may interchange l and d. One may think of much of this work as assigning portions of the brain to different subroutines in a program.

Writing depends, at least in less sophisticated subjects, upon a preliminary phonemic analysis with phonemes being recoded into graphemes in such a way as to preserve the order of sound and letter elements; while arithmetic and geometric operations have a different psychological structure. Each of these forms of activity is, Luria feels, based on the interaction of strictly defined zones of cerebral cortex--and he proposes this as the foundation for neuropsychological diagnosis of local brain lesions.

The accomplishment of a given activity may be impossible in some circumstances and simple in other for a patient with a local brain lesion--as the patient who after many unsuccessful attempts to say "No" said "No, doctor, I am absolutely unable to say 'no'", or the patient who cannot touch his nose on request, but can dislodge a fly from it.

The continuation of this work (see, e.g., A. R. Luria, E. G. Simernitskaya and B. Tubylevich [1970]) suggests that, as a result of frequent repetition, psychological operations not only become more rapid and cease to require special conscious control, but change their psychological structure and come to be effected through the activity of a different system of cortical zones. Here, by way of evidence, let us note Luria's finding that a local brain lesion disrupting the normal course of a certain psychological process may leave unaffected the course of more consolidated automatized operations, which are thus presumably based on quite a different cerebral control system. The following studies of defects in writing correlated with local brain lesions support the case:

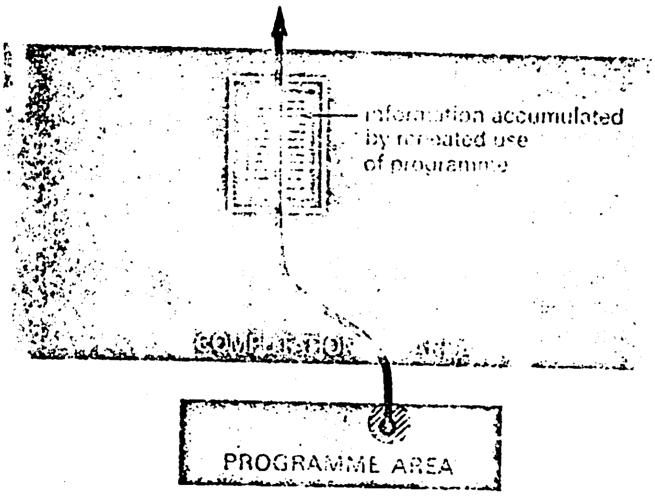
One woman with a vascular lesion of the speech zones of the left hemisphere was unable to write a single word from dictation, but could easily cope with the same task when asked to write "quickly, without a moment's thought". Apparently this avoided acoustic analysis (which was disturbed in this patient) and allowed the expression of the intact kinesthetic stereotype. A woman who had (probably bilateral) lesions involving predominantly the parieto-occipital region could neither copy letters, nor write them from dictation, but was able to write these letters if they were included in whole well-assimilated words. She could also write the alphabet correctly. A man with a large abscess in the left parietal lobe close to the mid-line experienced great difficulties when asked to write a word slowly, consciously analyzing its components--sometimes changing spatial location of elements, e.g., mirror or inverted disposition--but could often succeed when asked to write a word rapidly, especially when embedded in a whole phrase.

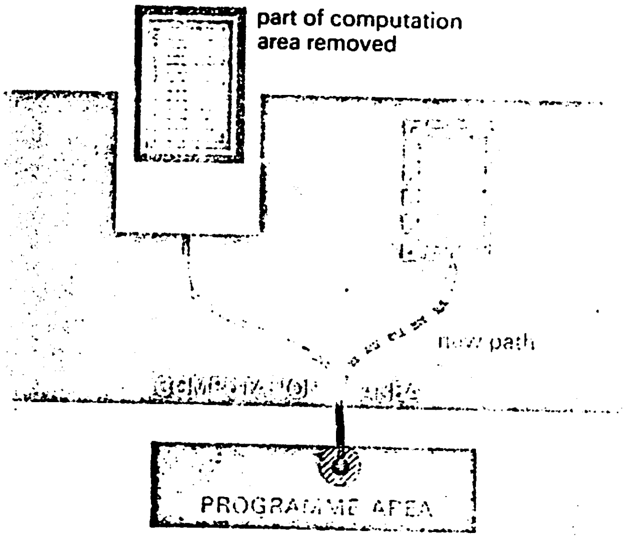
The clinic offers other related data: patients who are unable to repeat a word may utter it easily in the context of a well-assimilated phrase; some patients who cannot read relatively rare words can recognize "by sight" habitual "ideograms". Recognition of familiar words may remain intact even in cases where the reading of unfamiliar words, or writing them from dictation, is markedly deranged. One might explain this on the basis of frequent use yielding a multiplicity of "traces", which are thus less sensitive to damage.

In our computer jargon, we might say that the above patients behaved as if they could no longer go from the name of a subroutine to its entry point, but could use the subroutine in those programs which already included the entry point, rather than requiring explicit generation of the entry point anew each time the routine was required.

4.d. DIPM: A Distributed Information Processing Machine.

In trying to conceptualize the effects of brain damage, I have developed a theoretical model of a distributed information processing machine (DIPM) as a new mode of computer organization with resistance to certain forms of damage. In some central area of the machine are stored programs to execute which the computer must assign space from a surrounding computation area. If the program is used at all frequently, a specific area may be used by the program again and again. However, should the specified computation area be damaged, mechanisms still exist for switching in a new area. The logic functions of the machine are thus distributed and not centralized. In the computation areas new inputs actively interact with instructions and previously accumulated results. Thus no single center exists in which damage could lead to the breakdown of the whole machine--each fault leads to no worse than a partial impairment.





By removing or ablating parts of the brain of rats, and studying the effect on the animal's performance, K. S. Lashley found--with gross measurements of performance of simple tasks--that all areas of rat cortex seem equally important to the animal's ability to run through a maze; and that deterioration in the rat's performance is proportional to the amount of brain tissue removed. Of course, such laws are highly task specific and so, if a DIPM model is to guide the study of ablation experiments, we must use a series of interconnected DIPMs to model a brain which is executing tasks which require action linked to several different sensory modalities. In any case, as we have already said, if an ablation leaves the program itself undamaged, the program has to switch in a new computing area, and then resume its action. Thus if only a small computing region has been removed, this will cause little trouble. However, if the area removed is large there may be major problems of competition for the remaining space, as well as routing problems caused by the severing of major pathways carrying messages in and out of the relevant computation area. In short, the impairment may be relatively non-specific, but will increase as the size of the area removed increases. At the other extreme, if an area is removed which contains the DIPM's sole copy of a program, then that function will be lost completely, save to the extent that the machine can approximate it with other functions.

In repeated use of a program, the machine will store in its computation area much useful material which will substantially ease later computation. If such an area is then removed, performance will be greatly impaired, even were the program to be undamaged and the switching monitor to immediately assign a satisfactory new computation area. As repeated use gradually re-accumulated the lost material, so performance would return to the level prior to damage.

This models Luria's thesis that it is indeed highly probable that a psychological operation changes not only its structure, but also its cerebral organization, in the course of its functional development. Crude though the DIPM model is, it does suggest that such phenomena fit well into our cybernetic framework. In explaining how the brain is able to shift function in response to childhood damage one can build on the DIPM model, but a careful balance of "neural specificity" on the one hand and "sufficient structure to adapt" on the other seems required.

5. CONCLUSION

(Points of Intersection of Computer Science and Linguistics:
Artificial Intelligence as a proving ground for "real" semantics;
Language as a "distant horizon" for detailed brain theory;
a "current frontier" for studies of gross brain organization.)

In real problem-solving, changing the representation or viewpoint is an integral part of searching for a solution. Most artificial intelligence work has centered on searching for a solution within a given framework. What this boils down to is that the solution of a problem may be easy if you have a language which encompasses two different styles of discourse, so that methods in one domain may be applied to problems in another. Perhaps the most intriguing problem for the student of language from the cybernetic viewpoint is to understand how we may go back and forth between the vague hierarchical distributed internal model of the world embodied in a neural network, and the rather precise structures we can fashion out of words, which can be built up to great logical depth, and so be used to drive the internal model to go far beyond what would otherwise be the limits of perception.