An earlier version appeared in:
The analysis of Meaning: Informatics 5
Proceedings ASLIB/BCS conference, Oxford, March 1979,
Eds: M.MacCafferty and K.Gray, Published by Aslib.

THE PRIMACY OF NON-COMMUNICATIVE LANGUAGE

Aaron Sloman
School of Computer Science
The University of Birmingham
Birmingham
B15 2TT, England
http://www.cs.bham.ac.uk/~axs/

(Written in 1979 while at The University of Sussex)

Introduction

How is it possible for symbols to be used to refer to or describe things? I shall approach this question indirectly by criticising a collection of widely held views of which the central one is that meaning is essentially concerned with communication. A consequence of this view is that anything which could be reasonably described as a language is essentially concerned with communication. I shall try to show that widely known facts, for instance facts about the behaviour of animals, and facts about human language learning and use, suggest that this belief, and closely related assumptions (see A1 to A3, below) are false. Support for an alternative framework of assumptions is beginning to emerge from work in Artificial Intelligence, work concerned not only with language but also with perception, learning, problem-solving and other mental processes. The subject has not yet matured sufficiently for the new paradigm to be clearly articulated. The aim of this paper is to help to formulate a new framework of assumptions, synthesising ideas from Artificial Intelligence and Philosophy of Science and Mathematics. The rival frameworks can be briefly over-stated thus:

> OLD: A language is essentially a *social* phenomenon and meanings are essentially things to be communicated, so that it is impossible for anything to use a language solely for private purposes: *the primacy of communication.*
>
> NEW: the essence of language is storage of information for use and manipulation by an individual, and communicative potential is an evolutionary side-effect of this function: *the primacy of representation.*

A very clear formulation of the first thesis can be found in chapter 2 of Lyons (1977), and a not so clear but very influential discussion in Wittgenstein (1953). John Lyons has drawn my attention to Chomsky (1975), which criticises versions of the first thesis propounded by Searle, Grice and Strawson. The work of formal semanticists and mathematical linguists is usually neutral on this issue, since they are not concerned to explain how it is possible to use language meaningfully, but merely explore consequences of formal assumptions about meaning. Something like the second thesis is implicit or explicit in a great deal of work in Artificial Intelligence over the last twenty years or so (for surveys see Boden 1977, Winston 1977) and a related version is expounded at length in Fodor (1976). My own (1978) takes the second thesis for granted throughout. As should become clear later, the second thesis does not deny that there are languages (e.g. English, French...) which are used largely for communication, nor that many of their main features derive from this use. The claim is that use for communication with other individuals is not a necessary pre-condition

for the meaningful use of some language.

The standard framework

The old thesis is part of a collection of widely held assumptions which I shall challenge. Here is a summary of a central subset.

A1. The primary function of language is communication between individuals (e.g. Lyons writes: 'it is difficult to imagine any satisfactory definition of the term 'language' that did not incorporate some reference to the notion of communication' (1977), page 32, and goes on to state that it is obvious that it is impossible to account for meaning except in terms of communication). So language is essentially social.

A2. Learning a public, shared, language is a pre-condition of having knowledge, beliefs, intentions, principles, and of thinking, deciding or inferring.

A3. Human beings are the only animals which use language to describe things and reason with.

These assumptions are not necessarily all held simultaneously, for they are independent of one another (though I shall not try to prove that now). But they are often held together. I shall argue against them all, trying in particular to show that there are at least three senses in which the use of a rich and powerful internal language *within* an individual is prior to the use of language in overt communication between individuals. I then sketch a theory of how it is possible for an internal language to be used to refer to and describe an external world.

I do not claim that there is any *one* inner language common to all animals or even all human beings (compare Fodor 1976), for the inner language or languages of any one individual would develop under the influence of that individual's unique history, including possibly events prior to birth and certainly after exposure to a public language. It is a consequence of the theory sketched below that a language may be extended by the addition of new symbols not defined in terms of the previously known symbols. It follows that some learning of an overt language may *extend* an individual's inner language, rather than always simply relying on a fixed inner language to define new symbols. It could even be argued that the evolution of social systems, using shared overt languages, would, through natural selection, influence the genetically determined internal linguistic abilities, for instance allowing them to be more open to external influence, and thereby facilitating cultural evolution, which permits more rapid adaptation to changing circumstances than Darwinian evolution. Thus some of the innate linguistic abilities of a social animal might be geared to the communicative function of language. But I shall try to show that this is not a necessary condition for having linguistic abilities.

What is a language?

Of course, one can easily define "language" in such a way as to restrict language to overt communication between individuals, and that would make A1, above, true, but true by definition and misleading, as I shall show. In any case, to talk of individuals communicating by means of symbols implies that they can *understand* the symbols, i.e. interpret them as meaningful, and how this is possible cannot be explained without reference to internal processes. Stipulative definitions don't help with this.

There is a clear and important sense of the word "language" which does not make A1 true by definition -- and in this interpretation A1 is still widely assumed to be true. What is this broader sense of "language"? I shall give only an incomplete answer. I believe that most students of language would, at least after some reflection, accept the following as *necessary* conditions for saying that X uses a language L, even if they are not *sufficient* conditions. (The first three conditions derive from the work of Frege, but are now widely accepted.)

L1. L includes both simple and complex symbols, the latter being composed of the former, in a principled fashion. (Symbols are any kind of entity used in constructing maps, descriptions, representations etc. Non-denoting symbols, like parentheses and other syntactic devices may be included.)

L2. There is at any one time a definite set of *simple symbols* of L known to X (or usable by X), although this set may be enlarged over time, and some of the symbols may fall out of use.

L3. There is at any one time a restricted set of *modes of composition* of more complex symbols of L from less complex ones known to X, though the set of rules of composition (grammatical rules) may change, and need not be explicitly formulated in X or anywhere else: e.g. they may be consequences of other features of the procedures employed by X for using the symbols.

L4. X does not merely construct or contemplate such symbols, but uses them at least (a) to express beliefs and possible beliefs i.e. to represent what is or may be the case (b) to formulate questions i.e. to specify missing information (c) to formulate goals or purposes or intentions, or instructions. We need not assume that these different uses correspond to different subsets of L. For instance it may be that which of these uses X makes of a particular symbol varies from context to context, or even that no symbol is ever used with exactly one of these functions.

I shall not now attempt to define precisely what is meant by such words as "symbol", "rule", "facts", "questions","goals", "instructions", etc., or to go into all the many and subtle distinctions made by logicians and linguists and discussed at length in Lyons 1977 Vol 1. I assume everyone has at least a rough and ready grasp of L1 to L4, and can make some sense of the distinctions in L4 between using symbols to record what is the case, using them to specify gaps in what is recorded as being the case, and using them to generate behaviour by describing the behaviour. (How these things can be done is another matter.) Partial analyses will be offered later. It is not implied by L1 to L4 that X need be conscious of using L.

L1 to L4 do not say anything explicitly about communication between individuals. So it is not obviously true by definition that something which is a language in the sense implicitly defined by L1-L4 is primarily used, or even used at all, for communication between individuals. X might use L entirely in a private diary, or in its mind only, as far as L1 to L4 are concerned. For instance, X may formulate a question specifying missing information in the course of constructing a plan -- the question may be used to generate inferences and information-gathering processes. It need not be addressed to another individual. Similarly, instructions may be part of a stored plan or strategy used by X.

Nevertheless, I think it is widely held even if only implicitly that in some sense the main or primary use of anything which would be a language in the sense of L1-L4 must necessarily be overt communication between individuals. Other, more private, uses would have to be derivative, in some sense. The most powerful exponent of this essentially public view of language was Wittgenstein (1953). However, related, though less sophisticated, views are quite common. What is wrong with this cluster of views? Once again my approach will be indirect.

Intelligence in lesser mortals

Have you ever wondered how it is possible for animals to learn circus tricks? Or how birds manage to build nests? Or how it is possible for monkeys to leap through trees at high-speed without frequently crashing into branches or missing them completely and falling to the ground? Or how hunting animals find their way back to their lairs? Or how frogs, flies, and other apparently stupid animals are able to manoeuvre themselves into the right position to mate with other individuals? Or how a new-born deer can run after its mother? Or how spiders manage to make their webs in a variety of geometrically different physical situations, and to patch them when they are damaged?

Of course, it is possible to assume that such things just happen, that they are "natural", that no explanation is required, just as it is possible not to be puzzled about the fact that unsupported apples move towards not away from the earth, or the fact that frogs' eggs eventually grow into frogs and not fish. This flabby acceptance of facts may suffice for phenomenologically minded philosophers and the man in the street, but if one wishes to understand the possibility of such phenomena it is necessary to attempt to construct theories about *underlying mechanisms.*

At present no adequate theories about underlying mechanisms are available for the abilities listed in the previous paragraph. There are many *theory-building tools,* including concepts of physics and analogies with physical processes, concepts of neuro-physiology, concepts and formalisms of control-theory and systems-theory, and most recently concepts and formalisms of computing science and artificial intelligence. The latter are concerned with mechanisms which generate processes in which symbols are constructed and manipulated. At the moment it looks as if only this last set of *computational* theory-building tools has any hope of being useful for building theories about mechanisms which could generate both the variety and the fine-structure of the intelligent behaviour of animals.

Claim:

> No non-computational mechanism currently known is capable of generating a range of qualitiatively different patterns of behaviour intricately related and adapted to both the sensed structure of the environment and to pre-existing goals.

What, then, is a computational mechanism?

The general form of Artificial Intelligence Theories

Water running down a hill will, to a certain extent, avoid obstacles. But there is no need to assume that it has the goal of getting to the bottom of the hill, or that any intelligence is involved in generating its behaviour. Each portion of water merely responds in accordance with relatively simple mechanical principles to *local* conditions, and the overall behaviour is simply the *sum* of all these local processes. Thus something like the mathematics of differential equations and boundary conditions, possibly enhanced by singularity theory, suffices to represent and explain what is going on - even though in many cases measuring the boundary conditions and solving the equations may present very great technical difficulties.

In particular, there is no need to assume that the water makes use of a *representation* of the current situation which is compared with a *representation* of a goal situation, or that the results of such a comparison lead to the selection or construction of some strategy whose execution requires the collaboration of sub-systems which are under the control of a central executive. Processes like these would require a computational, i.e. symbol manipulating, mechanism.

By contrast, theories in A.I. are concerned with mechanisms which build, compare, manipulate, search for, interpret, analyse, or obey symbolic structures of some kind. The existing theories have many limitations, such as a lack of parallelism, a restriction to discrete (digital) symbolisms, and, above all, a very small amount of information (compared with what a human or animal brain seems able to store). Moreover, AI programs so far have had a very simple structure: for example there are none which could be described as even approximately like a complete organism with its own system of goals, perceptual abilities, planning abilities, and learning abilities. However, some of these restrictions are beginning to be overcome, and others probably will be, as far as can be judged at present.

Types of symbol manipulating mechanism

What I am claiming then is that the only paradigm of theory construction which looks remotely like being able to provide theories accounting for much animal behaviour is the computational paradigm, which describes mechanisms using internal symbolisms. We can distinguish different degrees and kinds of sophistication in such mechanisms. The following is but a short list of examples:

1. A single branch-free program is used, which, once triggered, always causes essentially the same sequence of instructions to be obeyed.

2. Whilst a program is being obeyed, sense-organs are continually updating some symbol store, and at certain points conditional instructions generate behaviour which depends on this incoming information, e.g. adjusting muscular exertion to the wind resistance.

3. As in the previous case, except that incoming information affects not just local behaviour (i.e. what is done at particular steps), but the global flow of control, as in a program which under certain conditions will transfer control to a quite different program e.g. a switch from food-gathering behaviour to escaping behaviour triggered by the smell of a predator, for instance.

4. Alongside other behaviour there may be a process of analysis of incoming information, producing an internal symbolic representation of the current environment, whether or not it is relevant to current needs and strategies: e.g. the construction of descriptions of relatively static three dimensional objects and relationships on the basis of continually varying two-dimensional retinal information, or the construction of some kind of map of the environment on the basis of exploring a sequence of routes through it. (Must one of these evolve before the other? Both require the ability to represent spatial information.)

5. Instead of using a permanent set of stored programs the organism may modify its programs, or synthesise new ones, in the light of an analysis of the short-comings of the old ones. This presupposes internal descriptions of some of the programs and of both their intended and their actual effects.

6. Instead of having a permanent set of stored procedures for making major decisions, on the basis of available information, the system may include procedures for modifying its decision-making strategies, including both the alteration of relative weightings of previously used criteria, and the synthesis of new principles and policies.

This is not meant to be anything like an exhaustive survey of types of symbol-using mechanisms which might be offered as explanations of increasingly sophisticated patterns of animal behaviour. The examples do, however, illustrate a number of dimensions in which computational systems can vary, namely:

A: the extent to which decisions (including decisions about how to make decisions, etc.) are postponed till "run-time"

B: the extent to which information is taken in and stored *in case* it may be useful, reducing the reliance on the immediate environment to provide information for decision-making processes.

C: the extent to which the system alters, or synthesises, its own programs.

D: the extent to which different activities can go on in parallel, e.g. performing actions, monitoring their effects, taking in new information, reconsidering goals and plans, etc.

Variations in these dimensions would account for variations in degrees of flexibility, generality of learning abilities, ability to solve problems, ability to adapt to changing circumstances, etc. The evolution of consciousness is probably connected with diversification of functions alluded to in D.

Whether or not mechanisms of the general forms sketched above do underlie the intelligent behaviour of animals, it is at least clear from work in computing science and artificial intelligence that such mechanisms *can exist*, and that they are capable, in principle, of generating many kinds of behaviour (internal and external) previously thought to be restricted to humans, since previous concepts of "mechanism" were based on analogies with relatively simple physical systems, like clocks, steam-engines, and telephone exchanges.

I have talked about mechanisms which use symbols, including symbols expressing instructions which can be obeyed, descriptions of aspects of the environment, and principles of decision-making. The most basic and primitive type of symbol-use is the execution of instructions. We can even treat the hill and the water flowing down it as a system in which the shape of the terrain amounts to a sort of stored symbolic program executed by the water under the influence of gravity and its internal constraints. But it is a very primitive kind of program, capable of generating a very limited class of behaviour, with little capacity for producing qualitatively varied behaviour in the light of information coming in from outside the system. An earthquake, or bomb, may change the program, but the program does not include tests explicitly anticipating changes, with alternative strategies for achieving goals. Nor can it cope with different goals at different times. Further, repeated execution may lead to changes, through soil erosion for example, but these can only be gradual and relatively continuous, unlike the sudden qualitative changes of behaviour of which a self-modifying computing system is capable. The system cannot hypothetically explore alternative internal changes then select one which fits some requirement.

This example is intended both to illustrate how broad the spectrum of mechanisms is which might be described as computational, and to illustrate that the kinds and degrees of difference between different locations on the spectrum may be so great that the metaphor of a *spectrum* is an oversimplification. The example also illustrates how the same chunk of reality may be viewed in different ways - e.g. as a physical system or as a computer executing instructions.

The semantics of internal languages

I have suggested that the most primitive and basic kind of symbol must be some kind of *instruction*, i.e. something which generates and controls behaviour in an appropriate *interpreter.* There is a very varied class of types of instruction, ranging from what might be thought of simply as physical causes (e.g. the shape of a hillside which only in a very extended sense can be said to instruct the water flowing down it), to very much more "descriptive" instructions which include a description of an action to be performed (e.g. "turn your head to the left"), or specify an end state to be achieved without specifying the action to achieve it (e.g. "Be here at noon tomorrow", or "find some food").

What we are beginning to understand, as a result of a series of increasingly complex computational experiments in the form of designing and implementing A.I. languages and programs, is that provided you have the first, most primitive, type of symbol-obeying system, in which the meaning of a symbol is little more than the effect it has on the machine (including such effects as changing some of the symbols), you can construct on top of it a series of layers of increasingly sophisticated virtual machines, including ones in which symbols are used to describe objects and their relationships, and eventually systems in which some of the symbols which are interpreted as instructions themselves contain *descriptive* elements: for instance in PLANNER-like languages where procedures are invoked not by name but by some kind of articulated pattern, which may function as a description of a state of affairs to be achieved. (E.g. see Winograd 1972).

> In short, *descriptive* meaning evolves out of *procedural* meaning, and more elaborate types of procedural meaning may evolve out of descriptive meaning. This evolution has occurred in computing science. Perhaps it also occurred in the development of living organisms.

We now return to the "central problem of semantics":

> Under what conditions can such a mechanism use some of its inner symbols as descriptions of what is the case, descriptions which, instead of merely producing some effect on the mechanism, refer to or describe things other than themselves, and do so correctly or incorrectly?

Note that the central idea is not having a meaning, but being used with a meaning. This central idea does not normally enter into formal theories of semantics such as the work of Tarksi: hence their limited interest for our purposes. I don't think the answer to the question is simple or obvious, and neither is it clear that existing computing systems have reached the kind of sophistication required for us to say that they *understand* symbols as descriptions, even though in many cases *we* clearly attach descriptive meaning to the symbols they use, including the internal data-structures. But we also attach significance to the contents of filing cabinets and tape recordings! A full discussion would require analysis of different kinds of referential and descriptive uses of symbols. A full analysis is not yet available. However, we can tentatively formulate some apparently necessary conditions for symbols to be used descriptively.

Preconditions for descriptive meaning

In order that a system S be said to use symbols from a language L to describe certain (types of) objects and their properties and relations, we could require the following conditions:

M1. S must be able to use sensory-detectors capable of receiving stimulation directly or indirectly (e.g. via light or sound waves) from the objects, the actual stimulation being determined in a principled fashion by the things and S's relationship to them (e.g. visual stimulation depends on viewpoint).

M2. S must be able both to *build* and to *reject* or *modify* descriptions using the language L, based on processes of analysis of the stimulation mentioned in M1.

M3. S must be able to make *inferences* from some descriptions formulated in L to others. That is to say, S must be able to use some descriptive symbolic structures as a starting point for building others related to them. (E.g. A.I. work on visual perception and the analysis of pictures shows how the construction of a 3-D interpretation involves an enormous amount of inference making)

M4. S must be capable of noting (in at least some cases) that two or more descriptions in L of some state of affairs cannot all be acceptable (i.e. they are inconsistent), and, in at least some cases, capable of taking steps to find out which should be rejected.

M5. S must be capable of using the descriptions in L as a basis for taking decisions about how to act. More precisely, S must be able to use some symbols as representations of possible states of affairs, and also be able to build a description of a series of possible actions which would make such a state of affairs actual.

M6. S must be capable of discovering (whether or not it expresses the discovery in L) that it lacks some information and using a complex symbol in L to specify what is missing and guide a process of attempting to acquire the information either by inference from other available descriptions, or by using the sense organs. I.e. S can use some symbols of L as questions.

These conditions will be familiar to philosophers of language. They will be relaxed somewhat later on. They do not completely define the semantic concepts they use, and even as incomplete definitions they are circular. It remains to be seen whether this circularity can be analysed as an acceptable case of mutual recursion. Even if the circularity is acceptable, the real work remains to be done, namely to flesh out these conditions for different kinds of symbols and different aspects of the world, including, for example, geometrical, physical, biological and social aspects of reality: the

preconditions for meaningfully using a word to refer to circles will be very different from the preconditions for meaningfully using a word to refer to cultural revolutions, for example.

Further analysis of M1 to M6 would take us into hoary debates about concept-empiricism, the verifiability and testability criteria of significance, distinctions between referring expressions and predicates, the role of quantifiers, modal operators, the importance of implicit definitions and meaning postulates, e.g. see Ayer (1946), Hempel (1950), Carnap (1956), Pap (1963), Popper (1959), Quine (1953) and many more. In particular, we should need to explain how a system may use symbols to describe objects, properties, and relationships, in a domain to which it has no direct access, so that it can never completely verify or falsify statements about the domain (see my 1978, chapter 9, and discussions by philosophers of science of the role of unobservables in theories, e.g. Pap (1963)).

An important idea in such philosophical debates is that implicit, partial, definitions (e.g. in the form of an axiom system) enable new concepts to get off the ground. For instance, a collection of axioms for Euclidean geometry in the context of a set of inference procedures, would partially and implicitly define concepts like "line", "point", "intersects", etc. In A.I. programs, e.g. programs concerned with describing visual scenes, instead of axioms and logical inference rules we often find a collection of procedures for building data-structures and for relating them to others. The procedures partially and implicitly define the meanings of the structures. This is a phenomenon crying out for more formal study.

The analogy with theoretical concepts of science and mathematics implies that not all newly-acquired concepts need be *translatable* into one's previous symbolism. (Compare Fodor 1976.) It also implies that the system may use predicates to describe the environment which are not definable explicitly in terms of tests which may be applied to sensory data. Instead, the descriptions are inferred from inconclusive tests on the basis of theoretical assumptions. For instance, the notion of a 'climbable object', or a 'surface moving nearer' need not be *defined* in terms of operations on retinal input, but may be *inferred* from descriptions of retinal input. The *meanings* of the symbols used to describe the environment, will be partially defined by the collection of inference rules (transformation and construction procedures) and theoretical postulates (inital data-structures) used. The postulates and inference rules need not take the forms studied by logicians: for instance, they may include the use of analogical representations in the sense defined in Sloman (1978), and many domain-specific inference procedures. The definitions implicit in such assumptions and procedures will be inherently incomplete, and the concepts indefinitely extendable by adding new theoretical assumptions about the nature of the reality referred to. These features are evident in theoretical concepts of science. It is not so easy to detect them in more familiar concepts like "hard", "wet", "dog", "food", etc., since we are less conscious of the inferences we make in ordinary life.

The essential incompleteness of semantics

Thus we may say that intelligent systems, like scientists, necessarily use symbols without full understanding, and without ever being able to establish finally whether what they say is true or false. But this is not something to lament: it is an inevitable fact about the semantics of a language used to represent information about things outside oneself. This fact seems to lie at the source of much philosophical discussion about knowledge and scepticism.

So the conditions M1 to M6 above do not imply that *every* descriptive or referential symbol S understands must be one which S can relate *directly*, using perceptual procedures, to the reality described or referred to. The symbol-system L may make contact with reality, e.g. through S's sense-organs, only at relatively scattered points, and only in indirect ways (like the connection between reality and our concepts of 'atom', 'gene', 'the distant past', 'the remote future', 'another person's mind', 'the cause of an event', 'Julius Caesar', 'the interior of the sun', 'the battle of

Hastings', and so on). The points of contact with reality may vary considerably from individual to individual, but this need not prevent different individuals storing much the same information about large chunks of the world. This is because their inference procedures permit them to extrapolate beyond what they have already learned. For instance because they can communicate, people who live in different places can share knowledge about the geography of the earth. And different animals who do not communicate may share knowledge about a forest, gleaned in diferent ways. All this has much in common with some views expounded in Quine 1953, and Strawson 1959.

For most of us, most of what we believe or think about is the result of a process of inference, hypothetical construction, use of indirect evidence, or acceptance of reports from intermediaries, but this doesn't stop us having beliefs and thoughts which refer to more or less remote portions of the world. The same could be true of other animals, or machines, even if their sources of information about the world are less rich, and include no other communicating intermediaries.

This view of the semantics of inner symbolisms implies that the inner language may be extended by the addition of new partially and implicitly defined symbols. Contrary to Fodor's claims, a new language may therefore be learnt without any new symbols being *translatable* into old ones -- more on this below. Hence different humans may use different "mentalese" even if they all started off the same. This admittedly sketchy analysis applies not just to the semantics of verbal or logical languages, but also to the use of maps and other analogical representations.

We could argue at length over whether *all* of the conditions M1 to M6 are necessary for S to use L with descriptive meaning, or whether some other necessary conditions should be added to the list, such as consciousness of the use being made of symbols. But such debates would be fruitless, amounting to little more than semantic squabbles over how we should use words like "symbol", "language", or "meaning". There are no doubt many different sorts of cases which could arise, forming yet another "spectrum" ranging from systems which satisfy only minimal conditions (see end of this paper) to systems as powerful as people. One of the goals of AI and Computing Science should be to explore this range of possibilities, using both theoretical analysis and computational experiments.

The primacy of inner languages

However, what is important in relation to assumption A1 is that the conditions M1 to M6 are intelligible, and that it makes sense to suppose that most or all of them might be satisfied by some symbol-using system which is not part of any society using any kind of overt language. Insofar as any communication is involved, it is only communication between sub-processes of a single system. Furthermore, I do not know how we can begin to explain the intelligence of many forms of animals without assuming that they make use of such internal symbol-systems. The rich variety of behaviour, the extent to which they can match fine details of their behaviour to the requirements of the environment, the ability to generalise from one situation to others, the apparent ability to acquire information and then use it on another slightly different occasion, e.g. to avoid danger or to find a new way home -- all these seem incapable of being explained without reference to internal processes in which information is stored in some symbol system.

Furthermore, facts about human infants and the work on learning in A.I. (e.g. Winston 1975, Sussman 1975) strongly suggest that human learning, including early language-learning and the development of sensori-motor skills, could not occur without the prior existence in infants of rich and complex symbol-manipulating systems capable of forming, testing, and modifying both plans and theories. My efforts to find alternative explanations in the writings of developmental psychologists, such as Piaget, have unearthed only vague hand-waving, or metaphorical re-description of observed behaviour. Work on AI systems which process English and other natural languages suggests that even the use of an *overt* language requires the use of internal symbolisms

for building up descriptions and interpretations of fragments of sentences, and for making inferences from what is actually said.

All this implies that there are at least three senses in which the use of an internal symbolism with descriptive and procedural semantics is prior to, or more fundamental than, the use of an overt language for communication between individuals.

P1. The use of some kinds of inner languages must have evolved before the evolution of what we normally call language, since intelligent animals existed before social languages. (Note that I am assuming that something not too different from Darwin's theory of biological evolution is correct. Theists may reach different conclusions.)

P2. The use of an inner language is a precondition for the learning of a human language like English or Urdu.

P3. The use of an inner language is a precondition for the continued use of external languages, but the converse does not hold (in view of P1 and P2).

I summarise all this in the slogan: *representation (or symbolisation) is prior to communication.*

If all this is correct then the three assumptions A1 to A3 are false. To rescue the assumption that language essentially involves communication by making it true by stipulation, would conceal important facts about the possibility of internal symbolisms which share several functions with external languages.

Wittgenstein's private language argument

The theory sketched here may appear to fall foul of Wittgenstein's (1953) arguments against the possibility of a private language. His argument is that the notion of following a rule is inapplicable to the use of some "logically private" symbolism since the correct/incorrect distinction could not be used when there is no possible public check that the rule has or has not been followed. This argument has a very dubious status, but, as Fodor remarks, it is irrelevant to computational theories, since nothing said above implies that the inner symbolism is *logically* inaccessible to outside scrutiny. In practice the difficulty of opening up a brain, or computer and working out what is going on may be insurmountable, but that is another matter.

How should meaning be represented?

It is important to distinguish two questions

(a) How should a theorist (e.g. linguist, psychologist, logician), represent the meanings of symbols of certain kinds?

(b) How are the meanings of the symbols represented by their user?

The answers to these questions may be the same for some users of some symbols, but they need not be. When a computer "understands" some machine language, it does not have or use any explicit representation of the meaning. Rather the existence of built-in machinery for interpreting (obeying) instructions gives the symbolism its meaning. This need not prevent computer scientists from attempting to describe the semantics of the language explicitly.

Similarly if a relatively high-level language is interpreted, instead of being compiled, (a possibility Fodor never discusses) then the stored symbols may be capable of generating behaviour in a systematic fashion but there need not be any separate internal representation of their meaning: what meaning they have is implicitly assigned by the procedures for interpreting them -- possibly in a context-sensitive fashion. The same applies to descriptive (non-procedural) stored symbolism, which is implicitly defined by the way the system is used, as outlined above. This is discussed

briefly by Fodor in connection with "meaning postulates" (1976, pp.149,ff), but only in relation to the learned public language. He never considers that the "mentalese" symbolism may, at least in part, be assigned a meaning in just this way. The native mentalese can then be gradually extended by the addition of new stored representations, partially and implicitly defining some new primitive symbols, and modifying the meanings (use) of old ones. (It might also evolve through development of the underlying interpreter.) The upshot would be a system in which there is no clear functional distinction between the concepts of mentalese and those of some other language. This is the sort of thing which justifies a claim to have learnt to think in a new language. Fodor's analogy with compiling a high-level programming language breaks down. However, this is still too vague and leaves open a wide range of possibilties for the internal symbolism - including stored sentences in an essentially public language (as when we memorise a poem or a set of directions to get to the station), "analogical" representations - e.g. 2-D arrays depicting retinal images or maps, or lists of ordered items representing the order of events or objects in the world (e.g. a memorised list of names of winners at Wimbledon, or a network of routes) and no doubt many more.

Reflecting consciously on the meaning of an English sentence can tell us little about the myriad unconscious processes involved in producing it, understanding it, inferring it, or believing what it says. We still need to learn a great deal about the trade-offs involved in alternative types of symbolisation. Working in the A.I. paradigm forces one to address issues about memory, about problem-solving, about recognition, about learning, about ways of achieving efficiency. For instance, work in A.I. suggests that if decisions and interpretations are required quickly, it will often be useful to store information in a highly redundant form. Peano's axioms may suffice for a logician interested in number theory, but no computational system frequently having to solve arithmetical problems could do so reasonably quickly without storing large numbers of 'partial results', which are logically redundant. Otherwise enormous searches among possible derivations from the axioms would be required for each new arithmetical task. Similarly, the use of an economically represented generative grammar might be much more time-consuming than the use of a far more redundant system, including what Becker (1975) describes as a 'phrasal lexicon'. This redundant system would be especially useful if different rules could be processed in parallel, and if incoming information was often incomplete or degraded by noise, mis-pronunciation, slips of the tongue, sloppy sentence-construction, etc. Items rejected by the 'basic' rules might be found to match relatively large stored schemas quite well. Once redundancy enters into the system, the scope for inconsistency increases. This could be a major factor in the development of language. Thus the criteria favoured by mathematical linguists, such as economy and consistency of grammars, might be the last things we should require of either efficient working systems or theories of how people and animals use their private and public languages. We need to explore these and other issues by designing and analysing new symbol-using systems displaying various forms of intelligence. The subject is so young, there are bound to be many surprises in store for us.

Solipsistic Intelligence

One of the surprises may be that we have to weaken some of the conditions listed above for using symbols with descriptive meaning, such as M1 and M5. There could be two machines running programs P1 and P2, the former connected to TV cameras and mechanical arms, as well as a teletype, and the latter only to a teletype. If P1 satisfies the conditions given above, and P2 is a subset of P1, then under certain circumstances we may be able to say that P2 contains all of P1 except the links to TV cameras, and perhaps the software for some of the lowest level analysis of sensory input. Thus if P1 can learn about the world either through its cameras, or through the teletype, then why should we deny that P2 can learn about the world through the teletype alone, like a blind and paralysed person who does not lack the computational abilities underlying sight and physical motion? P2 will not acquire so much information so easily. It may have to spend much effort speculating about details P1 perceives, and it might generate and accept more false

hypotheses.

> Thus, using symbols to formulate beliefs and hypotheses about an external world does not require that the world in fact be sensed and acted on, only that the internal symbols and procedures be sufficiently rich to have the potential to support such processes of interaction with the world.

As far as I know, A.I. work on language understanding has not yet really begun to address the question of how such potential can be shown to exist in a program which can communicate only via a teletype. It may be that there is no adequate test short of actually embedding the program in a more complete system, though I hope it will turn out that theoretical analysis will be possible instead. The thought processes of such a system might not be too different from some human thought processes disconnected from the real world, such as religious and metaphysical thinking, and some kinds of mathematical thinking, for instance about infinite-dimensional spaces. The final step is to notice that not even the teletype is necessary!

Acknowledgements

Discussions with Frank O'Gorman, Phil Johnson-Laird, Steve Draper, Bill Woods, Laurie Hollings, John Lyons and students attending graduate seminars in the Cognitive Studies Programme at Sussex University, helped me to formulate some of the issues and revise an early draft, as did unpublished work by Gerald Gazdar on 'Constituent Structures'. Some of the problems arose out of work on a project on "Computational flexibility in visual perception", funded by the SRC grant BRG/8688.7. Judith Dennison helped with production.

BIBLIOGRAPHY

Ayer, A.J., *Language Truth and Logic* 2nd edition, Gollancz, 1946.

Becker,, J.D. 'The phrasal lexicon', in R. Schank and B. Nash-Webber (eds) *Theoretical Issues in Natural Language Processing* Association of Computational Linguistics, 1975.

Boden, Margaret, *Artificial Intelligence and Natural Man* Harvester Press, and Basic Books. 1977.

Carnap, R., *Meaning and Necessity* Phoenix Books 1956.

Chomsky, Noam, *Reflections on Language* Temple Smith, and Fontana, 1976.

Fodor, J.A., *The Language of Thought* Harvester Press 1976.

Hempel, C.G, 'The Empiricist Criterion of Meaning' in A.J. Ayer (Ed.) Logical Positivism, The Free Press, 1959. Originally in *Revue Int. de Philosophie,_Vol.4.* 1950.

Lyons, John, *Semantics* Cambridge University Press. 1977.

Pap, A., *An Introduction to the Philosophy of Science* Eyre and Spottiswoode (Chapters 2-3). 1963.

Popper, K.R., *The Logic of Scientific Discovery* Hutchinson, 1959

Quine, W.V.O., 'Two Dogmas of Empiricism' in *From a Logical point of view* 1953.

Sloman, Aaron, *The Computer Revolution in Philosophy:* Philosophy Science and Models of Mind, Harvester Press and The Humanities Press. 1978.

Strawson, P. F., *Individuals: An Essay in Descriptive Metaphysics,* Methuen. 1959.

Sussman, G.J., *A Computer Model of Skill Acquisition* American Elsevier. 1975.

Winograd, T., *Understanding Natural Language* Edinburgh University Press 1972.

Winston, P.H., *The Psychology of Computer Vision* McGraw Hill, 1975.

Winston, P.H., *Artificial Intelligence,* Addison Wesley, 1977.

Wittgenstein, L., *Philosophical Investigations,* Blackwell, 1953.