To appear in *Action To Language via the Mirror Neuron System* (Michael A. Arbib, Editor), Cambridge University Press, 2005.

The Origin and Evolution of Language: A Plausible, Strong-AI Account

Jerry R. Hobbs

USC Information Sciences Institute

Marina del Rey, California

ABSTRACT

A large part of the mystery of the origin of language is the difficulty we experience in trying to imagine what the intermediate stages along the way to language could have been. An elegant, detailed, formal account of how discourse interpretation works in terms of a mode of inference called abduction, or inference to the best explanation, enables us to spell out with some precision a quite plausible sequence of such stages. In this chapter I outline plausible sequences for two of the key features of language – Gricean nonnatural meaning and syntax. I then speculate on the time in the evolution of modern humans each of these steps may have occurred.

1 FRAMEWORK

In this chapter I show in outline how human language as we know it could have evolved incrementally from mental capacities it is reasonable to attribute to lower primates and other mammals. I do so within the framework of a formal computational theory of language understanding (Hobbs et al., 1993). In the first section I describe some of the key elements in the theory, especially as it relates to the evolution of linguistic capabilities. In the next two sections I describe plausible incremental paths to two key aspects of language – -- meaning and syntax. In the final section I discuss various considerations of the time course of these processes.

1.1. Strong AI

It is desirable for psychology to provide a reduction in principle of intelligent, or intentional, behavior to neurophysiology. Because of the extreme complexity of the human brain, more than the sketchiest account is not likely to be possible in the near future. Nevertheless, the central metaphor of cognitive science, "The brain is a computer", gives us hope. Prior to the computer metaphor, we had no idea of what could possibly be the bridge between beliefs and ion transport. Now we have an idea. In the long history of inquiry into the nature of mind, the computer metaphor gives us, for the first time, the promise of linking the entities and processes of intentional psychology to the underlying biological processes of neurons, and hence to physical processes. We could say that the computer metaphor is the first, best hope of materialism.

The jump between neurophysiology and intentional psychology is a huge one. We are more likely to succeed in linking the two if we can identify some intermediate levels. A view that is popular these days identifies two intermediate levels – the symbolic and the connectionist.

Intentional Level

|

Symbolic Level

|

Connectionist Level

|

Neurophysiological Level

The intentional level is implemented in the symbolic level, which is implemented in the connectionist level, which is implemented in the neurophysiological level.[1] From the "strong AI" perspective, the aim of cognitive science is to show how entities and processes at each level emerge from the entities and processes of the level below. The reasons for this strategy are clear. We can observe intelligent activity and we can observe the firing of neurons, but there is no obvious way of linking these two together. So we decompose the problem into three smaller problems. We can formulate theories at the symbolic level that can, at least in a small way so far, explain some aspects of intelligent behavior; here we work from intelligent activity down. We can formulate theories at the connectionist level in terms of elements that are a simplified model of what we know of the neuron's behavior; here we work from the neuron up. Finally, efforts are being made to implement the key elements of symbolic processing in connectionist architecture. If each of these three efforts were to succeed, we would have the whole picture.

---

[1]Variations on this view dispense with the symbolic or with the connectionist level.

In my view, this picture looks very promising indeed. Mainstream AI and cognitive science have taken it to be their task to show how intentional phenomena can be implemented by symbolic processes. The elements in a connectionist network are modeled on certain properties of neurons. The principal problems in linking the symbolic and connectionist levels are representing predicate-argument relations in connectionist networks, implementing variable-binding or universal instantiation in connectionist networks, and defining the right notion of “defeasibility” or “nonmonotonicity” in logic[2] to reflect the “soft corners” that make connectionist models so attractive. Progress is being made on all these problems (e.g., Shastri and Ajjanagade, 1993; Shastri, 1999).

Although we do not know how each of these levels is implemented in the level below, nor indeed *whether* it is, we know that it *could* be, and that at least is something.

1.2. Logic as the Language of Thought

A very large body of work in AI begins with the assumptions that information and knowledge should be represented in first-order logic and that reasoning is theorem-proving. On the face of it, this seems implausible as a model for people. It certainly doesn't seem as if we are using logic when we are thinking, and if we are, why are so many of our thoughts and actions so illogical? In fact, there are psychological experiments that purport to show that people do not use logic in thinking about a problem (e.g., Wason and Johnson-Laird, 1972).

I believe that the claim that logic is the language of thought comes to less than one might think, however, and that thus it is more controversial than it ought to be. It is the claim that a broad range of cognitive processes are amenable to a high-level description in which six key features are present. The first three of these features characterize propositional logic and the next two first-order logic. I will express them in terms of “concepts”, but one can just as easily substitute propositions, neural elements, or a number of other terms.

- Conjunction: There is an additive effect ($P \wedge Q$) of two distinct concepts (P and Q) being activated at the same time.
- Modus Ponens: The activation of one concept (P) triggers the activation of another concept (Q) because of the existence of some structural relation between them ($P \supset Q$).
- Recognition of Obvious Contradictions: The recognition of contradictions in general is undecidable, but we have no trouble with the easy ones, for example, that cats aren't dogs.

---

[2] See Section 1.2.

- Predicate-Argument Relations: Concepts can be related to other concepts in several different ways. We can distinguish between a dog biting a man (*bite*(*D,M*)) and a man biting a dog (*bite*(*M,D*)).
- Universal Instantiation (or Variable Binding): We can keep separate our knowledge of general (universal) principles ("All men are mortal") and our knowledge of their instantiations for particular individuals ("Socrates is a man" and "Socrates is mortal").

Any plausible proposal for a language of thought must have at least these features, and once you have these features you have first-order logic. Note that in this list there are no complex rules for double negations or for contrapositives (if P implies Q then not Q implies not P). In fact, most of the psychological experiments purporting to show that people don't use logic really show that they don't use the contrapositive rule or that they don't handle double negations well. If the tasks in those experiments were recast into problems involving the use of modus ponens, no one would think to do the experiments because it is obvious that people would have no trouble with the task.

There is one further property we need of the logic if we are to use it for representing and reasoning about commonsense world knowledge -- defeasibility or nonmonotonicity. Our knowledge is not certain. Different proofs of the same fact may have different consequences, and one proof can be "better" than another.

The mode of defeasible reasoning used here is "abduction"[3], or inference to the best explanation. Briefly, one tries to prove something, but where there is insufficient knowledge, one can make assumptions. One proof is better than another if it makes fewer, more plausible assumptions, and if the knowledge it uses is more plausible and more salient. This is spelled out in detail in Hobbs et al. (1993). The key idea is that intelligent agents understand their environment by coming up with the best underlying explanations for the observables in it. Generally not everything required for the explanation is known, and assumptions have to be made. Typically, abductive proofs have the following structure.

We want to prove R.

We know $P \wedge Q \supset R$.

We know P.

We assume Q.

---

[3] A term due to Pierce (1955 [1903]).

We conclude *R*.

A logic is "monotonic" if once we conclude something, it will always be true. Abduction is "nonmonotonic" because we could assume *Q* and thus conclude *R*, and later learn that *Q* is false.

There may be many *Q*'s that could be assumed to result in a proof (including *R* itself), giving us alternative possible proofs, and thus alternative possible and possibly mutually inconsistent explanations or interpretations. So we need a kind of "cost function" for selecting the best proof. Among the factors that will make one proof better than another are the shortness of the proof, the plausibility and salience of the axioms used, a smaller number of assumptions, and the exploitation of the natural redundancy of discourse. A more complete description of the cost function is found in Hobbs et al. (1993).

1.3. Discourse Interpretation: Examples of Definite Reference

In the "Interpretation as Abduction" framework, world knowledge is expressed as defeasible logical axioms. To interpret the content of a discourse is to find the best explanation for it, that is, to find a minimal-cost abductive proof of its logical form. To interpret a sentence is to deduce its syntactic structure and hence its logical form, and simultaneously to prove that logical form abductively. To interpret suprasentential discourse is to interpret individual segments, down to the sentential level, and to abduce relations among them.

Consider as an example the problem of resolving definite references. The following four examples are sometimes taken to illustrate four different kinds of definite reference.

I bought a new car last week. *The car* is already giving me trouble.

I bought a new car last week. *The vehicle* is already giving me trouble.

I bought a new car last week. *The engine* is already giving me trouble.

The engine of my new car is already giving me trouble.

In the first example, the same word is used in the definite noun phrase as in its antecedent. In the second example, a hyponym is used. In the third example, the reference is not to the "antecedent" but to an object that is related to it, requiring what Clark (1975) called a "bridging inference". The fourth example is a determinative definite noun phrase, rather than an anaphoric one; all the information required for its resolution is found in the noun phrase itself.

These distinctions are insignificant in the abductive approach. In each case we need to prove the existence of the definite entity. In the first example it is immediate. In the second, we use the axiom

$(\forall\, x)\ car(x) \supset vehicle(x)$

In the third example, we use the axiom

$(\forall\, x)\ car(x) \supset (\exists\, y)\ engine(y,x)$

that is, cars have engines. In the fourth example, we use the same axiom, but after assuming the existence of the speaker's new car.

This last axiom is "defeasible" since it is not always true; some cars don't have engines. To indicate this formally in the abduction framework, we can add another proposition to the antecedent of this rule.

$(\forall\, x)\ car(x) \wedge etc_i(x) \supset (\exists\, y)\ engine(y,x)$

The proposition $etc_i(x)$ means something like "and other unspecified properties of *x*". This particular *etc* predicate would appear in no other axioms, and thus it could never be proved. But it could be assumed, at a cost, and could thus be a part of the least-cost abductive proof of the content of the sentence. This maneuver implements defeasibility in a set of first-order logical axioms operated on by an abductive theorem prover.

1.4. Syntax in the Abduction Framework

Syntax can be integrated into this framework in a thorough fashion, as described at length in Hobbs (1998). In this treatment, the predication

(1) $Syn\ (w,e,\ldots)$

says that the string *w* is a grammatical, interpretable string of words describing the situation or entity *e*. For example, *Syn*("John reads *Hamlet*", *e,…*) says that the string "John reads *Hamlet*." (*w*) describes the event *e* (the reading by John of the play *Hamlet*). The arguments of *Syn* indicated by the dots include information about complements and various agreement features.

Composition is effected by axioms of the form

(2) $Syn(w_1,\ e,\ \ldots,\ y,\ \ldots) \wedge Syn(w_2,\ y,\ \ldots) \supset Syn(w_1w_2,\ e,\ \ldots)$

A string w_1 whose head describes the eventuality *e* and which is missing an argument *y* can be concatenated with a string w_2 describing *y*, yielding a string describing *e*. For example, the string "reads" (w_1), describing a reading event *e* but missing the object *y* of the reading, can be concatenated with the string "*Hamlet*" (w_2) describing a book

y, to yield a string “reads *Hamlet*” (w_1w_2), giving a richer description of the event e in that it does not lack the object of the reading.

The interface between syntax and world knowledge is effected by “lexical axioms” of a form illustrated by

(3) $read'(e,x,y) \wedge text(y) \supset Syn(\text{“read”}, e, ..., x, ..., y, \ldots)$

This says that if e is the eventuality of x reading y (the logical form fragment supplied by the word “read”), where y is a text (the selectional constraint imposed by the verb “read” on its object), then e can be described by a phrase headed by the word “read” provided it picks up, as subject and object, phrases of the right sort describing x and y.

To interpret a sentence w, one seeks to show it is a grammatical, interpretable string of words by proving there in an eventuality e that it describes, that is, by proving (1). One does so by decomposing it via composition axioms like (2) and bottoming out in lexical axioms like (3). This yields the logical form of the sentence, which then must be proved abductively, the characterization of interpretation we gave in Section 1.3.

A substantial fragment of English grammar is cast into this framework in Hobbs (1998), which closely follows Pollard and Sag (1994).

1.5 Discourse Structure

When confronting an entire coherent discourse by one or more speakers, one must break it into interpretable segments and show that those segments themselves are coherently related. That is, one must use a rule like

$Segment(w_1, e_1) \wedge Segment(w_2, e_2) \wedge rel(e,e_1,e_2) \supset Segment(w_1w_2, e)$

That is, if w_1 and w_2 are interpretable segments describing situations e_1 and e_2 respectively, and e_1 and e_2 stand in some relation *rel* to each other, then the concatenation of w_1 and w_2 constitutes an interpretable segment, describing a situation e that is determined by the relation. More about the possible relations in Section 4.

This rule applies recursively and bottoms out in sentences.

$Syn(w, e, \ldots) \supset Segment(w, e)$

A grammatical, interpretable sentence w describing eventuality e is a coherent segment of discourse describing e. This axiom effects the interface between syntax and discourse structure. *Syn* is the predicate whose axioms characterize syntactic structure; *Segment* is the predicate whose axioms characterize discourse structure; and they meet in this axiom. The predicate *Segment* says that string w is a *coherent* description of an eventuality e; the predicate *Syn* says that string *w* is a *grammatical and interpretable* description of eventuality *e*; and this axiom says that being grammatical and interpretable is one way of being coherent.

To interpret a discourse, we break it into coherently related successively smaller segments until we reach the level of sentences. Then we do a syntactic analysis of the sentences, bottoming out in their logical form, which we then prove abductively.[4]

1.6 Discourse as a Purposeful Activity

This view of discourse interpretation is embedded in a view of interpretation in general in which an agent, to interpret the environment, must find the best explanation for the observables in that environment, which includes other agents.

An intelligent agent is embedded in the world and must, at each instant, understand the current situation. The agent does so by finding an explanation for what is perceived. Put differently, the agent must explain why the complete set of observables encountered constitutes a coherent situation. Other agents in the environment are viewed as intentional, that is, as planning mechanisms, and this means that the best explanation of their observable actions is most likely to be that the actions are steps in a coherent plan. Thus, making sense of an environment that includes other agents entails making sense of the other agents' actions in terms of what they are intended to achieve. When those actions are utterances, the utterances must be understood as actions in a plan the agents are trying to effect. The speaker's plan must be recognized.

Generally, when a speaker says something it is with the goal that the hearer believe the content of the utterance, or think about it, or consider it, or take some other cognitive stance toward it. Let us subsume all these mental terms under the term “cognize”. We can then say that to interpret a speaker *A*'s utterance to *B* of some content, we must explain the following:

goal(*A, cognize*(*B,* content-of-discourse)

Interpreting the content of the discourse is what we described above. In addition to this, one must explain in what way it serves the goals of the speaker to change the mental state of the hearer to include some mental stance toward the content of the discourse. We must fit the act of uttering that content into the speaker's presumed plan.

---

[4] This is an idealized, after-the-fact picture of the result of the process. In fact, interpretation, or the building up of this structure, proceeds word-by-word as we hear or read the discourse.

The defeasible axiom that encapsulates this is

$(\forall\, s, h, e_1, e, w)[goal(s, e_1) \wedge cognize'(e_1, h, e) \wedge Segment(w, e) \supset utter(s, h, w)]$

That is, normally if a speaker *s* has a goal e_1 of the hearer *h* cognizing a situation *e* and *w* is a string of words that conveys *e*, then *s* will utter *w* to *h*. It is only defeasible because there are multiple strings *w* that can convey *e*.

We appeal to this axiom to interpret the utterance as an intentional communicative act. That is, if *A* utters to *B* a string of words *W*, then to explain this observable event, we have to prove *utter*(*A,B,W*). That is, just as interpreting an observed flash of light is finding an explanation for it, interpreting an observed utterance of a string *W* by one person *A* to another person *B* is to find an explanation for it. We begin to do this by backchaining on the above axiom. Reasoning about the speaker's plan is a matter of establishing the first two propositions in the antecedent of the axiom. Determining the informational content of the utterance is a matter of establishing the third. The two sides of the proof influence each other since they share variables and since a minimal proof will result when both are explained and when their explanations use much of the same knowledge.

1.7 A Structured Connectionist Realization of Abduction

Because of its elegance and very broad coverage, the abduction model is very appealing on the symbolic level. But to be a plausible candidate for how people understand language, there must be an account of how it could be implemented in neurons. In fact, the abduction framework can be realized in a structured connectionist model called SHRUTI developed by Lokendra Shastri (Shastri and Ajjanagadde, 1993; Shastri, 1999). The key idea is that nodes representing the same variable fire in synchrony. Substantial work must be done in neurophysics to determine whether this kind of model is what actually exists in the human brain, although there is suggestive evidence. A good recent review of the evidence for the binding-via-synchrony hypothesis is given in Engel and Singer (2001). A related article by Fell et al. (2001) reports results on gamma band synchronization and desynchronization between parahippocampal regions and the hippocampus proper during episodic memory memorization.

By linking the symbolic and connectionist levels, one at least provides a proof of *possibility* for the abductive framework.

There is a range of connectionist models. Among those that try to capture logical structure in the structure of the network, there has been good success in implementing defeasible *propositional* logic. Indeed, nearly all the applications to natural language processing in this tradition begin by setting up the problem so that it is a problem in propositional logic. But this is not adequate for natural language understanding in general. For example, the coreference problem, e.g., resolving pronouns to their antecedents, requires the expressivity of first-order logic even to state. We need a way of expressing predicate-argument relations and a way of expressing different instantiations of the same general principle. We need a mechanism for universal instantiation, that is, the binding of variables to specific entities. In the connectionist literature, this has gone under the name of the *variable-binding problem*.

The essential idea behind the SHRUTI architecture is simple and elegant. A predication is represented as an assemblage or cluster of nodes, and axioms representing general knowledge are realized as connections among these clusters. Inference is accomplished by means of spreading activation through these structures.

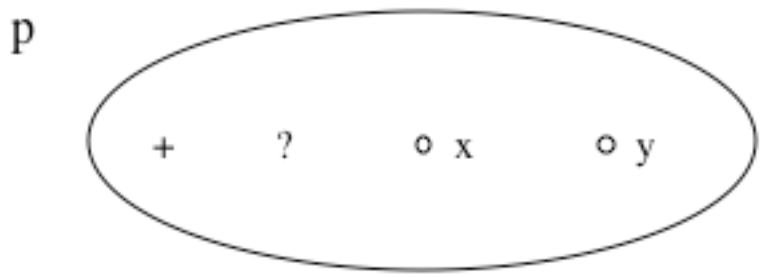

Figure 1 Predicate cluster for $p(x,y)$. The collector node (+) fires asynchronously in proportion to how plausible it is that $p(x,y)$ is part of the desired proof. The enabler node (?) fires asynchronously in proportion to how much $p(x,y)$ is required in the proof. The argument nodes for x and y fire in synchrony with argument nodes in other predicate clusters that are bound to the same variable.

In the cluster representing predications (Figure 1), two nodes, a collector node and an enabler node, correspond to the predicate and fire asynchronously. That is, they don't need to fire synchronously, in contrast to the "argument nodes" described below; for the collector and enabler nodes, only the *level* of activation matters. The level of activation on the enabler node keeps track of the "utility" of this predication in the proof that is being searched for. That is, the activation is higher the greater the need to find a proof for this predication, and thus the more expensive it is to assume. The level of activation on the collector node is higher the greater the plausibility that this predication is part of the desired proof.

We can think of the activations on the enabler nodes as prioritizing goal expressions, whereas the activations on the collector nodes indicate degree of belief in the predications, or more properly, degree of belief in the current relevance of the predications. The connections between nodes of different predication clusters have a strength of activation, or link weight, that corresponds to strength of association between the two concepts. This is one way we can capture the defeasibility of axioms in the SHRUTI model. The proof process then consists of activation spreading through enabler nodes, as we backchain through axioms, and spreading forward through collector nodes from something known or assumed. In addition, in the predication cluster, there are argument nodes, one for each argument of the predication. These fire synchronously with the argument nodes in other predication clusters to which they are connected. Thus, if the clusters for $p(x, y)$ and $q(z, x)$ are connected, with the two x nodes linked to each other, then the two x nodes will fire in synchrony, and the y and z nodes will fire at an offset with the x nodes and with each other. This synchronous firing indicates that the two x nodes represent variables bound to the same value. This constitutes the solution to the variable-binding problem. The role of variables in logic is to capture the identity of entities referred to in different places in a logical expression; in SHRUTI this identity is captured by the synchronous firing of linked nodes.

Proofs are searched for in parallel, and winner-takes-all circuitry suppresses all but the one whose collector nodes have the highest level of activation.

There are complications in this model for such things as managing different predications with the same predicate but different arguments. But the essential idea is as described. In brief, the view of relational information processing implied by SHRUTI is one where reasoning is a transient but systematic propagation of *rhythmic* activity over structured cell-ensembles, each active entity is a phase in the rhythmic activity, dynamic bindings are represented by the *synchronous* firing of appropriate nodes, long-term facts are circuits that detect coincidences in the ongoing flux of rhythmic activity, and rules are high-efficacy links that cause the propagation of rhythmic activity between cell-ensembles. Reasoning is the spontaneous outcome of a SHRUTI network.

In the abduction framework, the typical axiom in the knowledge base is of the form

(4) $(\forall x,y)[p_1(x,y) \wedge p_2(x,y) \supset (\exists z)[q_1(x,z) \wedge q_2(x,z)]]$

That is, the top-level logical connective will be implication. There may be multiple predications in the antecedent and in the consequent. There may be variables (x) that occur in both the antecedent and the consequent, variables (y) that occur only in the antecedent, and variables (z) that occur only in the consequent. Abduction backchains from predications in consequents of axioms to predications in antecedents. Every step in the search for a proof can be considered an abductive proof where all unproved predications are assumed for a cost. The best proof is the least cost proof.

The implementation of this axiom in SHRUTI requires predication clusters of nodes and axiom clusters of nodes (see Figure 1). A predication cluster, as described above, has one collector node and one enabler node, both firing asynchronously, corresponding to the predicate and one synchronously firing node for each argument. An axiom cluster has one collector node and one enabler node, both firing asynchronously, recording the plausibility and the utility, respectively, of this axiom participating in the best proof. It also has one synchronously firing node for each variable in the axiom -- in our example, nodes for x, y and z. The collector and enabler nodes fire asynchronously and what is significant is their level of activation or rate of firing. The argument nodes fire synchronously with other nodes, and what is significant is whether two nodes are the same or different in their phases.

The axiom is then encoded in a structure like that shown in Figure 2. There is a predication cluster for each of the predications in the axiom and one axiom cluster that links the predications of the consequent and antecedent. In general, the predication clusters will occur in many axioms; this is why their linkage in a particular axiom must be mediated by an axiom cluster.

Suppose (Figure 2) the proof process is backchaining from the predication $q_1(x,z)$. The activation on the enabler node (?) of the cluster for $q_1(x,z)$ induces an activation on the enabler node for the axiom cluster. This in turn induces activation on the predication nodes for $p_1(x,y)$ and $p_2(x,y)$. Meanwhile the firing of the x node in the q_1 cluster induces the x node of the axiom cluster to fire in synchrony with it, which in turn causes the x nodes of the p_1 and p_2 clusters to fire in synchrony as well. In addition, a link (not shown) from the enabler node of the axiom cluster to the y argument node of the same cluster causes the y argument node to fire, while links (not shown) from the x and z nodes cause that firing to be out of phase with the firing of the x and z nodes. This firing of the y node of the axiom cluster induces synchronous firing in the y nodes of the p_1 and p_2 clusters.

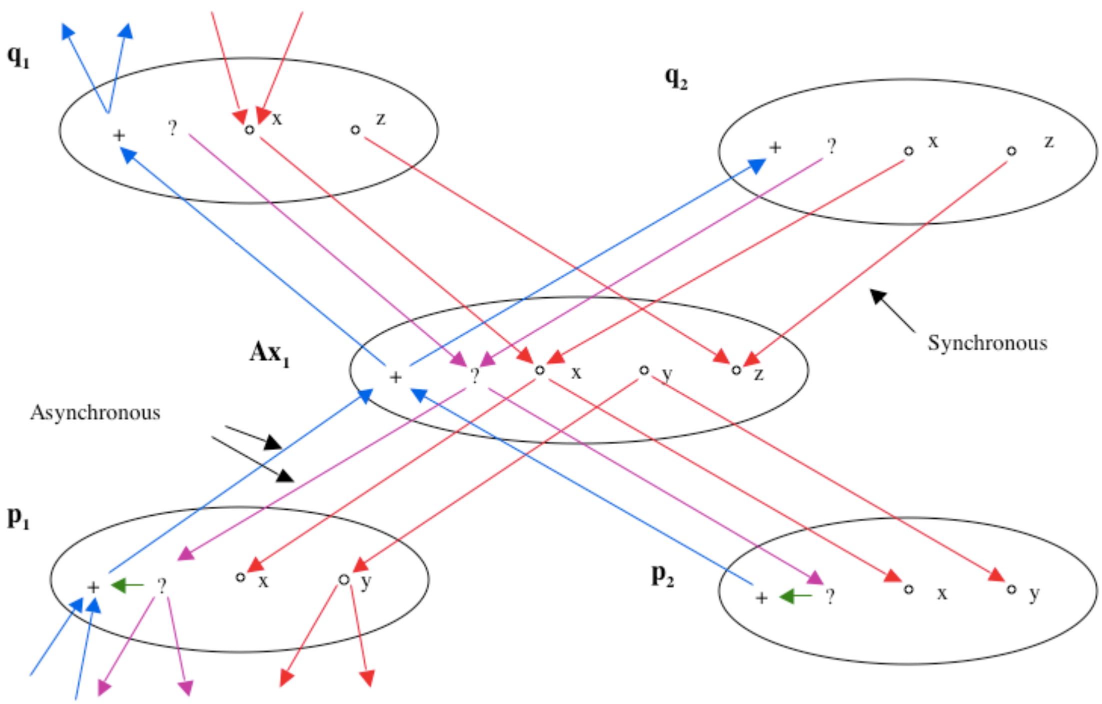

Figure 2 SHRUTI encoding of axiom $(\forall\ x,y)[p_1(x,y) \wedge p_2(x,y) \supset (\exists\ z)[q_1(x,z) \wedge q_2(x,z)]]$. Activation spreads backward from the enabler nodes (?) of the q_1 and q_2 clusters to that of the Ax_1 cluster and on to those of the p_1 and p_2 clusters, indicating the utility of this axiom in a possible proof. Activation spreads forward from the collector nodes (+) of the p_1 and p_2 clusters to that of the axiom cluster Ax_1 and on to those of the q_1 and q_2 clusters, indicating the plausibility of this axiom being used in the final proof. Links between the argument nodes cause them to fire in synchrony with other argument nodes representing the same variable.

By this means we have backchained over axiom (4) while keeping distinct the variables that are bound to different values. We are then ready to backchain over axioms in which p_1 and p_2 are in the consequent. As mentioned above, the q_1 cluster is linked to other axioms as well, and in the course of backchaining, it induces activation in those axioms' clusters too. In this way, the search for a proof proceeds in parallel. Inhibitory links suppress contradictory inferences and will eventually force a winner-takes-all outcome.

1.8 Incremental Changes to Axioms

In this framework, incremental increases in linguistic competence, and other knowledge as well, can be achieved by means of a small set of simple operations on the axioms in the knowledge base:

1. The introduction of a new predicate, where the utility of that predicate can be argued for cognition in general, independent of language.

2. The introduction of a new predicate p specializing an old predicate q:

$(\forall\, x)\ p(x) \supset q(x)$

For example, we learn that a beagle is a type of dog.

$(\forall\, x)\ beagle(x) \supset dog(x)$

3. The introduction of a new predicate p generalizing one or more old predicates q_i:

$(\forall\, x)\ q_1(x) \supset p(x),\ \ (\forall\, x)\ q_2(x) \supset p(x), \ldots$

For example, we learn that dogs and cats are both mammals.

$(\forall\, x)\ dog(x) \supset mammal(x),\ \ (\forall\, x)\ cat(x) \supset mammal(x)$

4. Increasing the arity of a predicate to allow more arguments.

$p(x) \Rightarrow p(x,y)$

For example, we learn that “mother” is not a property but a relation.

$mother(\mathrm{x}) \Rightarrow mother(x,y)$

5. Adding a proposition to the antecedent of an axiom.

$p_1(x) \supset q(x) \Rightarrow p_1(x) \wedge p_2(x) \supset q(x)$

For example, we might first believe that a seat is a chair, then learn that a seat with a back is a chair.

$seat(x) \supset chair(x) \Rightarrow seat(x) \wedge back(y,x) \supset chair(x)$

6. Adding a proposition to the consequent of an axiom.

$p(x) \supset q_1(x) \Rightarrow p(x) \supset q_1(x) \wedge q_2(x)$

For example, a child might see snow for the first time and see that it's white, and then goes outside and realizes it's also cold.

$snow(x) \supset white(x) \Rightarrow snow(x) \supset white(x) \wedge cold(x)$

All of these incremental changes can be implemented in the SHRUTI model via relatively simple means involving the recruitment of nodes, by strengthening latent connections as a response to frequent simultaneous activations (Shastri, 2001; Shastri and Wendelken, 2003; Wendelken and Shastri, 2003).

These incremental operations can be seen as constituting a plausible mechanism for both the development and the evolution of cognitive capacities. In this paper, I will show how the principal features of language could have resulted from a sequence of such incremental steps, starting from the cognitive capacity one could expect of ordinary primates.

1.9 Summary of Background

To summarize, the framework assumed in this chapter has the following features:

A detailed, plausible, computational model for a large range of linguistic behavior.

A possible implementation in a connectionist model.

An incremental model of learning, development, and evolution.

An implementation of that in terms of node recruitment.

In the remainder of the paper it is shown how two principal features of language – Gricean meaning and syntax – could have arisen from nonlinguistic cognition through the action of three mechanisms:

incremental changes to axioms,

folk theories required independent of language,

compilation of proofs into axioms.

These two features of language are, in a sense, *the* two key features of language. The first, Gricean meaning, tells how single words convey meaning in discourse. The second, syntax, tells how multiple words combine to convey complex meanings.

2 The Evolution of Gricean Meaning

In Gricean non-natural meaning, what is conveyed is not merely the content of the utterance, but also the intention of the speaker to convey that meaning, and the intention of the speaker to convey that meaning *by means of* that specific utterance. When A shouts “Fire!” to B, A expects that

1. B will believe there is a fire

2. B will believe A wants B to believe there is fire

3. 1 will happen because of 2

Five steps take us from natural meaning, as in "Smoke means fire," to Gricean meaning (Grice, 1948). Each step depends on certain background theories being in place, theories that are motivated even in the absence of language. Each new step in the progression introduces a new element of defeasibility. The steps are as follows:

1. Smoke means fire
2. "Fire!" means fire
3. Mediation by belief
4. Mediation by intention
5. Full Gricean meaning

Once we get into theories of belief and intention, there is very little that is certain. Thus, virtually all the axioms used in this section are defeasible. That is, they are true most of the time, and they often participate in the best explanation produced by abductive reasoning, but they are sometimes wrong. They are nevertheless useful to intelligent agents.

The theories that will be discussed in this section – belief, mutual belief, intention, and collective action – are some of the key elements of a theory of mind (Premack and Woodruff, 1978; Gordon, this volume). I discuss the possible courses of evolution of a theory of mind in Section 4.

2.1 Smoke Means Fire

The first required folk theory is a theory of causality (or rather, a number of theories *with* causality). There will be no definition of the predicate *cause*, that is, no set of necessary and sufficient conditions.

$cause(e_1, e_2) \equiv \ldots$

Rather there will be a number of domain-dependent theories saying what sorts of things cause what other sorts of things. There will be lots of necessary conditions

$cause(e_1, e_2) \supset \ldots$

and lots of sufficient conditions

$\ldots \supset cause(e_1,e_2)$

An example of the latter type of rule is

$smoke(y) \supset (\exists\, x)[fire(x) \wedge cause(x,y)]$

That is, if there's smoke, there's fire (that caused it).

This kind of causal knowledge enables prediction, and is required for the most rudimentary intelligent behavior.

Now suppose an agent B sees smoke. In the abductive account of intelligent behavior, an agent interprets the environment by telling the most plausible causal story. Here the story is that since fire causes smoke, there is a fire. B's seeing smoke causes B to believe there is fire, because B knows fire causes smoke.

2.2 "Fire!" Means Fire

Suppose seeing fire causes another agent A to emit a particular sound, say, "Fire!" and B knows this. Then we are in exactly the same situation as in Step 1. B's perceiving A making the sound "Fire!" causes B to believe there is a fire. B requires one new axiom about what causes what, but otherwise no new cognitive capabilities.

In this sense, sneezing means pollen, and "Ouch!" means pain. It has often been stated that one of the true innovations of language is its arbitrariness. The word "fire" is in no way iconic; its relation to fire is arbitrary and purely a matter of convention. The arbitrariness does not seem to me especially remarkable, however. A dog that has been trained to salivate when it hears a bell is responding to an association just as arbitrary as the relation between "fire" and fire.

I've analyzed this step in terms of comprehension, however, not production. Understanding a symbol-concept relation may require nothing more than causal associations. But producing a new symbol for a concept with the intention of using it for communication requires more in an underlying theory of mind. One at least has to have the notion of another individual's belief, since the aim of the new symbol is to create a belief in the other's mind.[5]

2.3 Mediation by Belief

For the next step we require a folk theory of belief, that is, a set of axioms explicating, though not necessarily defining, the predicate *believe*. The principal elements of a folk theory of belief are the following:

a. An event occurring in an agent's presence causes the agent to perceive the event.

   cause(*at*(*x, y, t*), *perceive*(*x, y, t*))[6]

This is only defeasible. Sometimes an individual doesn't know what's going on around him.

---

[5] I am indebted to George Miller and Michael Arbib for discussions on this point.

b. Perceiving an event causes the agent to believe the event occurred. (Seeing is believing.)

$cause(perceive(x, y, t), believe(x, y, t))$

c. Beliefs persist.

$t_1 < t_2 \supset cause(believe(x, y, t_1), believe(x, y, t_2))$

Again, this is defeasible, because people can change their minds and forget things.

d. Certain beliefs of an agent can cause certain actions by the agent. (This is an axiom schema, that can be instantiated in many ways.)

$cause(believe(x, P, t), ACT(x, t))$

For example, an individual may have the rule that an agent's believing there is fire causes the agent to utter "Fire!"

$fire(f) \supset cause(believe(x, f, t), utter(x, \text{“Fire!”}, t))$

Such a theory would be useful to an agent even in the absence of language, for it provides an explanation of how agents can transmit causality, that is, how an event can happen at one place and time and cause an action that happens at another place and time. It enables an individual to draw inferences about unseen events from the behavior of another individual. Belief functions as a carrier of information.

Such a theory of belief allows a more sophisticated interpretation, or explanation, of an agent A's utterance, "Fire!" A fire occurred in A's presence. Thus, A believed there was a fire. Thus, A uttered "Fire!" The link between the event and the utterance is mediated by belief. In particular, the observable event that needs to be explained is that an agent A uttered "Fire!" and the explanation is as follows:

$$utter(A, \text{“Fire!”}, t_2)$$

|

$$believe(A, f, t_2) \wedge fire(f)$$

|

$$believe(A, f, t_1) \wedge t_1 < t_2$$

|

[6] This is not the real notation because it embeds propositions within predicates, but it is more convenient for this chapter and conveys the essential meaning. An adequate logical notation for

$$perceive(a, f, t_1)$$

|

$$at(A, f, t_1)$$

There may well be other causes of a belief besides seeing. For example, communication with others might cause belief. Thus the above proof could have branched another way below the third line. This fact means that with this innovation, there is the possibility of "language" being cut loose from direct reference.

Jackendoff (1999) points out the distinction between two relics of one-word prelanguage in modern language. The word "ouch!", as pointed out above, falls under the case of Section 2.2; it is not necessarily communicative. The word "shh" by contrast has a necessary communicative function; it is uttered to induce a particular behavior on the part of the hearer. It requires that the speaker have some sort of theory of others' beliefs and how those beliefs are created and what behaviors they induce.

Note that this theory of belief could in principle be strictly a theory of other individuals, and not a theory of one's self. There is no need in this analysis that the interpreter even *have* a concept of self.

2.4 Near-Gricean Non-Natural Meaning

The next step is a close approximation of Gricean meaning. It requires a much richer cognitive model. In particular, three more background folk theories are needed, each again motivated independently of language. The first is a theory of goals, or intentionality. By adopting a theory that attributes agents' actions to their goals, one's ability to predict the actions of other agents is greatly enhanced. The principal elements of a theory of goals are the following:

a. If an agent x has an action by x as a goal, that will, defeasibly, cause x to perform this action. This is an axiom schema, instantiated for many different actions.

(5) $cause(goal(x,ACT(x)),ACT(x))$

That is, wanting to do something causes an agent to do it. Using this rule in reverse amounts to the attribution of intention. We see someone doing something and we assume they did it because they wanted to do it.

b. If an agent x has a goal g_1 and g_2 tends to cause g_1, then x may have g_2 as a goal.

(6) $cause(g_2, g_1) \supset cause(goal(x, g_1), goal(x, g_2))$

---

beliefs, causal relations, and so on can be found in Hobbs (1985a).

This is only a defeasible rule. There may be other ways to achieve the goal g_1, other than g_2. This rule corresponds to the body of a STRIPS planning operator as used in AI (Fikes and Nilsson, 1971). When we use this rule in the reverse direction, we are inferring an agent's ends from the means.

c. If an agent A has a goal g_1 and g_2 enables g_1, then A has g_2 as a goal.

(7) $enable(g_2,g_1) \supset cause(goal(x, g_1), goal(x, g_2))$

This rule corresponds to the prerequisites in the STRIPS planning operators of Fikes and Nilsson (1971).

Many actions are enabled by the agent knowing something. These are knowledge prerequisites. The form of these rules is

$enable(believe(x, P),ACT(x))$

The structure of goals linked in these ways constitutes a plan. To achieve a goal, one must make all the enabling conditions true and find an action that will cause the goal to be true, and do that.

The second required theory is a theory of joint action or collective intentionality. This is the same as a theory of individual intentionality, except that collectives of individuals can have goals and beliefs and can carry out actions. In addition, collective plans must bottom out in individual action. In particular, a group believes a proposition if every member of the group believes it. This is the point in the development of a theory of mind where a concept of self is probably required; one has to know that one is a member of the group like the rest of the community.

Agents can have as goals events that involve other agents. Thus, they can have in their plans knowledge prerequisites for other agents. A can have as a goal that B believe some fact. Communication is the satisfaction of such a goal.

The third theory is a theory of how agents understand. The essential content of this theory is that agents try to fit events into causal chains. The first rule is a kind of causal modus ponens. If an agent believes e_2 and believes e_2 causes e_3, that will cause the agent to believe e_3.

$cause(believe(x, e_2) \wedge believe(x, cause(e_2, e_3)), believe(x, e_3))$

This is defeasible since the individual may simply fail to draw the conclusion.

The second rule allows us to infer that agents backchain on enabling conditions. If an agent believes e_2 and believes e_1 enables e_2, then the agent will believe e_1

$cause(believe(x,e_2) \wedge believe(x, enable(e_1, e_2)), believe(x, e_1))$

The third rule allows us to infer that agents do causal abduction. That is, they look for causes of events that they know about. If an agent believes e_2 and believes e_1 causes e_2, then the agent may come to believe e_1.

$cause(believe(x, e_2) \wedge believe(x, cause(e_1, e_2)), believe(x, e_1))$

This is defeasible since the agent may have beliefs about other possible causes of e_2.

The final element of the folk theory of cognition is that all folk theories, including this one, are believed by every individual in the group. This is also defeasible. It is a corollary of this that A's uttering “Fire!” may cause B to believe there is a fire.

Now the near-Gricean explanation for the utterance is this: A uttered “Fire!” because A had the goal of uttering “Fire!”, because A had as a goal that B believe there is a fire, because B's belief is a knowledge prerequisite in some joint action that A has as a goal (perhaps merely joint survival) and because A believes there is a fire, because there was a fire in A's presence.

2.5 Full Gricean Non-Natural Meaning

Only one more step is needed for full Gricean meaning. It must be a part of B's explanation of A's utterance not only that A had as a goal that B believe there is a fire and that caused A to have the goal of uttering “Fire!”, but also that A had as a goal that A's uttering “Fire!” would cause B to believe there is a fire. To accomplish this we must split the planning axiom (6) into two:

(6a) If an agent A has a goal g_1 and g_2 tends to cause g_1, then A may have as a goal that g_2 cause g_1.

(6b) If an agent A has as a goal that g_2 cause g_1, then A has the goal g_2.

The planning axioms (5), (6), and (7) implement means-end analysis. This elaboration captures the intentionality of the means-end relation.

The capacity for language evolved over a long period of time, after and at the same time as a number of other cognitive capacities were evolving. Among the other capacities were theories of causality, belief, intention, understanding, joint action, and (nonlinguistic) communication. As the relevant elements of each of these capacities were acquired, they would have enabled the further development of language as well.

In Section 4 there is a discussion of possible evolutionary histories of these elements of a theory of mind.

3 THE EVOLUTION OF SYNTAX

3.1 The Two-Word Stage

When agents encounter two objects in the world that are adjacent, they need to explain this adjacency by finding a relation between the objects. Usually, the explanation for why something is where it is is that that is its normal place. It is normal to see a chair at a desk, and we don't ask for further explanation. But if something is out of place, we do. If we walk into a room and see a chair on a table, or we walk into a lecture hall and see a dog in the aisle, we wonder why.

Similarly, when agents hear two adjacent utterances, they need to explain the adjacency by finding a relation between them. A variety of relations are possible. “Mommy sock” might mean “This is Mommy's sock” and it might mean “Mommy, put my sock on”.

In general, the problem facing the agent can be characterized by the following pattern:

(8) $(\forall\, w_1,w_2,x,y,z)[B(w_1,y) \wedge C(w_2,z) \wedge rel(x,y,z) \supset A(w_1w_2,x)]$

That is, to recognize two adjacent words or strings of words w_1 and w_2 as a composite utterance of type A meaning x, one must recognize w_1 as an object of type B meaning y, recognize w_2 as an object of type C meaning z, and find some relation between y and z, where x is determined by the relation that is found. There will normally be multiple possible relations, but abduction will choose the best.

This is the characterization of what Bickerton (1990) calls “protolanguage”. One utters meaningful elements sequentially and the interpretation of the combination is determined by context. The utterance “Lion. Tree.” could mean there's a lion behind the tree or there's a lion nearby so let's climb that tree, or numerous other things. Bickerton gives several examples of protolanguage, including the language of children in the two-word phase and the language of apes. I'll offer another example: the language of panic. If a man runs out of his office shouting, “Help! Heart attack! John! My office! CPR! Just sitting there! 911! Help! Floor! Heart attack!” we don't need syntax to tell us that he was just sitting in his office with John when John had a heart attack, and John is now on the floor, and the man wants someone to call 911 and someone to apply CPR.

Most if not all rules of grammar can be seen as specializations and elaborations of pattern (8). The simplest example in English is compound nominals. To understand “turpentine jar” one must understand “turpentine” and "jar" and find the most plausible relation (in context) between turpentine and jars. In fact, compound nominals can be viewed as a relic of protolanguage in modern language.

Often with compound nominals the most plausible relation is a predicate-argument relation, where the head noun supplies the predicate and the prenominal noun supplies an argument. In "chemistry teacher", a teacher is a teacher of something, and the word "chemistry" tells us what that something is. In "language origin", something is originating, and the word "language" tells us what that something is.

The two-word utterance "Men work" can be viewed in the same way. We must find a relation between the two words to explain their adjacency. The relation we find is the predicate-argument relation, where "work" is the predicate and "men" is the argument.

The phrase structure rules

S → NP VP;   VP → V NP

can be written in the abductive framework (Hobbs, 1998) as

(9) $(\forall\, w_1,w_2,x,e)[Syn(w_1,x) \wedge Syn(w_2,e) \wedge Lsubj(x,e) \supset Syn(w_1w_2,e)]$

(10) $(\forall\, w_3,w_4,y,e)[Syn(w_3,e) \wedge Syn(w_4,y) \wedge Lobj(y,e) \supset Syn(w_3w_4,e)]$

In the first rule, if w_1 is string of words describing an entity x and w_2 is a string of words describing the eventuality e and x is the logical subject of e, then the concatenation w_1w_2 of the two strings can be used to describe e., in particular, a richer description of e specifying the logical subject. This means that to interpret w_1w_2 as describing some eventuality e, segment it into a string w_1 describing the logical subject of e and a string w_2 providing the rest of the information about e. The second rule is similar. These axioms instantiate pattern (8). The predicate *Syn*, which relates strings of words to the entities and situations they describe, plays the role of *A*, *B* and *C* in pattern (8), and the relation *rel* in pattern (8) is instantiated by the *Lsubj* and *Lobj* relations.

Syntax, at a first cut, can be viewed as a set of constraints on the interpretation of adjacency, specifically, as predicate-argument relations.

Rule (9) is not sufficiently constrained, since w_2 could already contain the subject. We can prevent this by adding to the arity of *Syn*, one of the incremental evolutionary modifications in rules in Section 1.8, and giving *Syn* a further argument indicating that something is missing.

(11) $(\forall\ w_1,w_2,x,e)[Syn(w_1,x,-,-) \wedge Syn(w_2,e,x,-) \wedge Lsubj(x,e) \supset Syn(w_1w_2,e,-,-)]$

(12) $(\forall\ w_3,w_4,y,e)[Syn(w_3,e,x,y) \wedge Syn(w_4,y,-,-) \wedge Lobj(y,e) \supset Syn(w_3w_4,e,x,-)]$

Now the expression $Syn(w_3, e, x, y)$ says something like "String w_3 would describe situation e if strings of words describing x and y can be found in the right places."

But when we restructure the axioms like this, the *Lsubj* and *Lobj* are no longer needed where they are, because the x and y arguments are now available at the lexical level. We can add axioms linking predicates in the knowledge base with words in the language. We then have following rules, where the lexical axiom is illustrative.

(13) $(\forall\ w_1,w_2,x,e)[Syn(w_1,x,-,-) \wedge Syn(w_2,e,x,-) \supset Syn(w_1w_2,e,-,-)]$

(14) $(\forall\ w_3,w_4,y,e)[Syn(w_3,e,x,y) \wedge Syn(w_4,y,-,-) \supset Syn(w_3w_4,e,x,-)]$

(15) $(\forall\ e,x,y)[read'(e,x,y) \wedge text(y) \supset Syn(\text{"read"}, e,x,y)]$

This is the form of the rules shown in Section 1.4. thus, the basic rules of syntax can be seen as direct developments from pattern (8). These rules describe the structure of syntactic knowledge; they do not presume any particular mode of processing that uses it.

We can add three more arguments to incorporate part-of-speech, agreement, and subcategorization constraints. As mentioned above, a rather extensive account of English syntax in this framework, similar to that in Pollard and Sag (1994), is given in Hobbs (1998).

Metonymy is a pervasive characteristic of discourse. When we say

I've read Shakespeare.

we coerce "Shakespeare" into something that can be read, namely, the writings of Shakespeare. So syntax is a set of constraints on the interpretation of adjacency as predicate-argument relations plus metonymy. Metonymy can be realized formally by the axiom

(16) $(\forall\ w,e,x,z)[Syn(w,e,x,-) \wedge rel(x,z) \supset Syn(w,e,z,-)]$

That is, if w is a string that would describe e providing the subject x of e is found, and *rel* is some metonymic "coercion" relation between x and z, then w can also be used as a string describing e if a subject describing z is found. Thus, z can stand in for x, as "Shakespeare" stands in for "the writings of Shakespeare". In this example, the metonymic relation *rel* would be *write*.

Metonymy is probably not a recent development in the evolution of language. Rather it is the most natural starting point for syntax. In many protolanguage utterances, the relation found between adjacent elements involves just such a metonymic coercion. Axiom (16) is a specialization of axiom (8), where it is two strings we are trying to relate and the relation is a composition of the predicate-argument relation (the writings are the logical object of the reading) and a metonymic relation (the writings were written by Shakespeare).

In multiword discourse, when a relation is found to link two words or larger segments into a composite unit, it too can be related to adjacent segments in various ways. The tree structure of sentences arises out of this recursion. Thus, "reads" and "*Hamlet*" concatenate into the segment "reads *Hamlet*", a verb phrase which can then concatenate with "John" to form the sentence "John reads *Hamlet*."

I have illustrated this advance – conveying predicate-argument relations by position – with the crucially important example of clause structure. But a similar story could be told about the equally important internal structure of noun phrases, which conveys a modification relation, a variety of the predicate-argument relation.

The competitive advantage this development confers is clear. There is less ambiguity in utterances and therefore more precision, and therefore more complex messages can be constructed. People can thereby engage in more complex joint action.

3.2 Signaling Predication and Modification

The languages of the world signal predication primarily by means of position and particles (or affixes). They signal modification primarily by means of adjacency and various concord phenomena. In what has been presented so far, we have seen how predicate-argument relatons can be recovered from adjacency. Japanese is a language that conveys predicate-argument relations primarily through postpositional particles, so it will be useful to show how this could have arisen by incremental changes from pattern (8) as well. For the purposes of this example, a simplified view of Japanese syntax is sufficient: A verb at the end of the sentence conveys the predicate; the Japanese verb "iki" conveys the predicate *go*. The verb is preceded by some number of postpositional phrases, in any order, where the noun phrase is the argument and the postposition indicates which argument it is; "kara" is a postposition meaning "from", so "Tokyo kara" conveys the information that Tokyo is the *from* argument of the verb.

Signaling predication by postpositions, as does Japanese, can be captured in axioms, specializing and elaborating pattern (8) and similar to (11), as follows:

$(\forall\, w_1,w_2,e,x)[Syn(w_1,x,\mathbf{n},-) \wedge Syn(w_2,e,\mathbf{p},x) \supset Syn(w_1w_2,e,\mathbf{p},-)]$

$(\forall\, w_3,w_4,e)[Syn(w_3,e,\mathbf{p},-) \wedge Syn(w_4,e,\mathbf{v},-) \supset Syn(w_3w_4,e,\mathbf{v},-)]$

$(\forall\, e,x)[from(e,x) \supset Syn(\text{"kara"},e,\mathbf{p},x)]$

$(\forall\, e)[go(e) \supset Syn(\text{"iki"},e,\mathbf{v},-)]$

The first rule combines a noun phrase and a postposition into a postpositional phrase. The second rule combines a postpositional phrase and a verb into a clause, and permits multiple postpositional phrases to be combined with the verb. The two lexical axioms link Japanese words with underlying world-knowledge predicates.[7] The fourth rule generates a logical form for the verb specifying the type of event it describes. The third rule links that event with the arguments described by the noun phrases via the relation specified by the postposition.

The other means of signaling predication and modification, such as inflection and agreement, can be represented similarly.

Klein and Perdue (1997), cited by Jackendoff (1999), identify features of what they call the Basic Variety in second-language learning, one of the most important of which is the Agent First word order; word order follows causal flow. Once means other than position are developed for signalling predicate-argument relations, various alternations are possible, including passives and the discontinuous elements discussed next, enabling language users to move beyond the Basic Variety.

3.3 Discontinuous Elements

Consider

John is likely to go.

To interpret this, an agent must find a relation between "John" and "is likely". Syntax says that it should be a predicate-argument relation plus a possible metonymy. The predicate "is likely" requires a proposition or eventuality as its argument, so we must coerce "John" into one. The next phrase "to go" provides the required metonymic coercion function. That John will go is likely. This analysis can be represented formally by the following axiom:

$(\forall\, w_3,w_4,e,e_1)[Syn(w_3,e,e_1,-) \wedge Syn(w_4, e_1,x,-) \supset Syn(w_3w_4,e,x,-)]$

John (x) stands in for John's going (e_1) where the relation between the two is provided by the phrase "to go" (w_2). This axiom has the form of axiom (16), where the x of (16) is e_1 here, the z of (16) is x here, and the $rel(x,z)$ of (16) is the $Syn(w_4, e_1,x,-)$ in this axiom. (Hobbs (2001) provides numerous examples of phenomena in English that can be analyzed in terms of interactions between syntax and metonymy.)

This locution is then reinterpreted as a modified form of the VP rule (14), by altering the first conjunct of the above axiom, giving us the VP rule for "raising" constructions.

$(\forall\, w_3,w_4,y,e)[Syn(w_3,e,x,e_1) \wedge Syn(w_4, e_1,x,-) \supset Syn(w_3w_4,e,x,-)]$

That is, if a string w_3 ("is likely") describing a situation e and looking for a logical subject referring to x (John) and a logical object referring to e_1 (John's going) is concatenated with a string w_2 ("to go") describing e_1 and looking for a subject x (John), then the result describes the situation e provided we can find a logical subject describing x.

This of course is only a plausible analysis of how discontinuous elements in the syntax could have arisen, but in my view the informal part of the analysis is very plausible, since it rests of the very pervasive interpretive move of metonymy. The formal part of the analysis is a direct translation of the informal part into the formal logical framework used here. When we do this translation, we see that the development is a matter of two simple incremental steps – specialization of a predicate (*rel* to *Syn*) and a modification of argument structure – that can be realized through the recruitment of nodes in a structured connectionist model.

3.4 Long-Distance Dependencies

One of the most "advanced" and probably one of the latest universal phenomena of language is long-distance dependencies, as illustrated by relative clauses and wh-questions. They are called *long-distance* dependencies because in principal the head noun can be an argument of a predication that is embedded arbitrarily deeply. In the noun phrase

the man John believes Mary said Bill saw

the man is the logical object of the seeing event, at the third level of embedding.

In accounting for the evolution of long-distance dependencies, we will take our cue from the Japanese. (For the purposes of this example, all one needs to know about Japanese syntax is that relative clauses have the form of clauses placed before the head noun.) It has been argued that the Japanese relative clause is as free as the English compound nominal in its interpretation. All that is required is that there be *some* relation between the situation

---

7 Apologies for using English as the language of thought in this example.

described by the relative clause and the entity described by the head noun (Akmajian and Kitagawa, 1974; Kameyama, 1994). They cite the following noun phrase as an example.

Hanako ga iede shita Taroo

Hanako Subj run-away-from-home did Taroo

Taroo such that Hanako ran away from home

Here it is up to the interpreter to find some plausible relation between Taroo and Hanako's running away from home.

We may take Japanese as an example of the basic case. Any relation will explain the adjacency of the relative clause and the noun. In English, a further constraint is added, analogous to the constraint between subject and verb. The relation must be the predicate-argument relation, where the head noun is the argument and the predicate is provided, roughly, by the top-level assertion in the relative clause and its successive clausal complements. Thus, in "the man who John saw", the relation between the man and the seeing event is the predicate-argument relation – the man is the logical object of the seeing. The clause "John saw ()" has a "gap" in it where the object should be, and that gap is filled by, loosely speaking, "the man". It is thus a specialization of pattern (8), and a constraint on the interpretation of the relation *rel* in pattern (8).

The constraints in French relative clauses lie somewhere between those of English and Japanese; it is much easier in French for the head to be an argument in an adjunct modifier of a noun in the relative clause. Other languages are more restrictive than English (Comrie, 1981, Chapter 7). Russian does not permit the head to be an argument of a clausal complement in the relative clause, and in Malagasy the head must be in subject position in the relative clause.

The English case can be incorporated into the grammar by increasing the arity of the *Syn* predicate, relating strings of words to their meaning. Before we had arguments for the string, the entity or situation it described, and the missing logical subject and object. We will increase the arity by one, and add an argument for the entity that will fill the gap in the relative clause. The rules for relative clauses then becomes

$(\forall\, w_1,e_1,x,y)[Syn(w_1,e_1,x,y,\text{-}) \wedge Syn(\text{``''},y,\text{-},\text{-},\text{-}) \supset Syn(w_1,e_1,x,\text{-},y)]$

$(\forall\, w_1,w_2,e_1,y)[Syn(w_1,y,\text{-},\text{-},\text{-}) \wedge Syn(w_2,e,\text{-},\text{-},y) \supset Syn(w_1w_2,y,\text{-},\text{-},\text{-})]$

The first rule introduces the gap. It says a string w_1 describing an eventuality e_1 looking for its logical object y can concatenate with the empty string provided the gap is eventually matched with a head describing y. The second rule says, roughly, that a head noun w_1 describing y can concatenate with a relative clause w_2 describing e but having a gap y to form a string $w_1 w_2$ that describes y. The rare reader interested in seeing the details of this treatment should consult Hobbs (1998).

In conversational English one sometimes hears "which" used as a subordinate conjunction, as in

I did terrible on that test, which I don't know if I can recover from it.

This can be seen as a relaxation of the constraint on English relative clauses, back to the protolanguage pattern of composition.

There are several ways of constructing relative clauses in the world's languages (Comrie, 1981, Chapter 7). Some languages, like Japanese, provide no information about which argument in the relative clause, if any, is identical to the head. But in all of the languages that do, this information can be captured in fairly simple axioms similar to those above for English. Essentially, the final argument of *Syn* indicates which argument is to be taken as coreferential with the head, however the language encodes that information.

Seeking a relation between adjacent or proximate words or larger segments in an utterance is simply an instance of seeking explanations for the observables in our environment, specifically, observable relations. Syntax can be seen largely as a set of constraints on such interpretations, primarily constraining the relation to the predicate-argument relation. The changes taking us from the protolanguage pattern (8) to these syntactic constraints are of three kinds, the first two of which we have discussed.

- Specializing predicates that characterize strings of words, as the predicate *Syn* specializes the predicates in pattern (8).
- Increasing the arity of the *Syn* predicate, i.e., adding arguments, to transmit arguments from one part of a sentence to another.
- Adding predications to antecedents of rules to capture agreement and subcategorization constraints.

The acquisition of syntax, whether in evolution or in development, can be seen as the accumulation of such constraints.

As mentioned above, the particular treatment of syntax used here closely follows that of Pollard and Sag (1994). They go to great efforts to show the equivalence of their Head-driven Phrase Structure Grammar to the Government and Binding theories of Chomsky (1981) then current, and out of which the more recent Minimalist theory of Chomsky (1995) has grown. It is often difficult for a computational linguist to see how Chomsky's theories could be realized computationally, and a corollary of that is that it is difficult to see how one could construct an incremental, computational account of the evolution of the linguistic capacity. By contrast, the unification grammar used by Pollard and Sag is transparently computational, and, as I have shown in this section, one can construct a compelling plausible story about the incremental development of the capacity for syntax. Because of the work Pollard and Sag have done in relating Chomsky's theories to their own, the account given in this chapter can be seen as a counterargument to a position that "Universal Grammar" had to have evolved as a whole, rather than incrementally. Jackendoff (1999) also presents compelling arguments for the incremental evolution of the language capacity, from a linguist's perspective.

4 Remarks on the Course of the Evolution of Language

Relevant dates in the time course of the evolution of language and language readiness are as follows:

1. Mammalian dominance: c65-50M years ago
2. Common ancestor of monkeys and great apes: c15M years ago
3. Common ancestor of hominids and chimpanzees: c5M years ago
4. Appearance of Homo erectus: c1.5M years ago
5. Appearance of Homo sapiens sapiens: c200-150K years ago
6. African/non-African split: c90K years ago
7. Appearance of preserved symbolic artifacts: c70-40K years ago
8. Time depth of language reconstruction: c10K years ago
9. Historical evidence: c5K years ago

In this section I will speculate about the times at which various components of language, as explicated in this chapter, evolved. I will then discuss two issues that have some prominence in this volume:

1. Was there a holophrastic stage before fully modern language? This is a question, probably, about the period just before the evolution of Homo sapiens sapiens.

2. When Homo sapiens sapiens evolved, did they have fully modern language or merely language readiness? This is a question about the period between the evolution of Homo sapiens sapiens and the appearance of preserved symbolic material culture.

4.1 Evolution of the Components of Language

Language is generally thought of as having three parts: phonology, syntax, and semantics. Language maps between sound (phonology) and meaning (semantics), and syntax provides the means for composing elementary mappings into complex mappings. The evolution of the components of language are illustrated in Figure 3.

According to Arbib (Chapter 1, this volume), gestural communication led to vocal communication, which is phonology. This arrow in the figure needs a bit of explication. Probably gesture and vocal communication have always both been there. It is very hard to imagine that there was a stage in hominid evolution when individuals sat quietly and communicated to each other by gesture, or a stage when they sat with their arms inert at their sides and chattered with each other. Each modality, as Goldstein, Byrd and Saltzman (this volume) point out, has its advantages. In some situations gestural communication would have been the most appropriate, and in others vocal communication. As Arbib points out, language developed in the region of the brain that had originally been associated with gesture and hand manipulations. In that sense gesture has a certain primacy. The most likely scenario is that there was a stage when manual gestures were the more expressive system. Articulatory gestures co-opted that region of the brain, and eventually became a more expressive system than the manual gestures.

In my view, the ability to understand composite event structure is a precursor to protolanguage, because the latter requires one to recover the relation between two elements. In protolanguage two-word utterances are composite events. Protolanguage led to syntax, as I argue in Section 3.1. For Arbib, composite event structure is part of the transition from protolanguage to language; for him this transition is post-biological. But he uses the term "protolanguage" to describe a holophrastic stage, which differs from the way I, and I think Bickerton, use the term.

According to the account in Section 2 of this chapter, the use of causal associations was the first requirement for the development of semantics. Causal association is possible in a brain that does the equivalent of propositional logic (such as most current neural models), but before one can have a theory of mind, one must have the equivalent of first-order logic. A creature must be able to distinguish different tokens of the same type. The last requirement is the development of a theory of mind, including models of belief, mutual belief, goals, and plans.

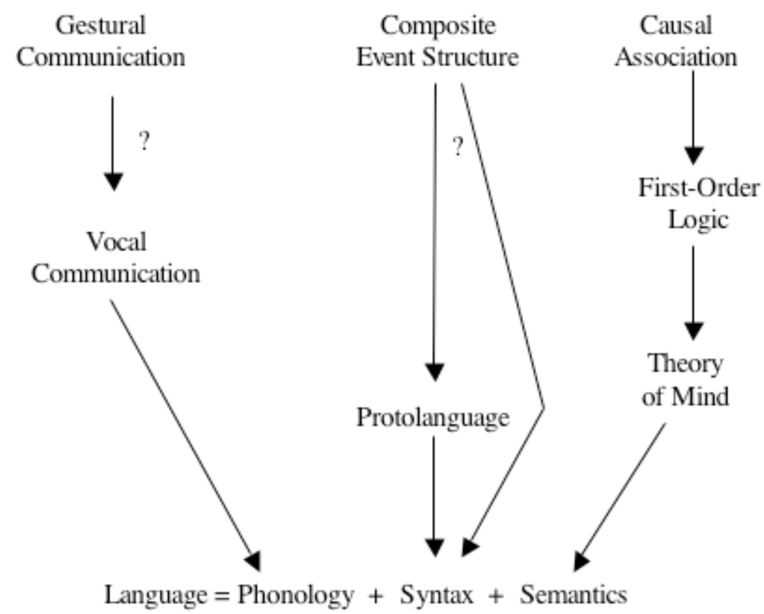

Figure 3: Evolution of the Components of Language

Arbib lists a number of features of language that have to evolve before we can say that fully modern language has evolved. It is useful to point out what elements in Figure 3 are necessary to support each of these features. *Naming*, or rather the interpretation of names, requires only causal association. A causal link must be set up between a sound and a physical entity in the world. Dogs and parrots can do this. Production in naming, by contrast, requires a theory of others' beliefs. *Parity* between comprehension and production requires a theory of the mind of others. The *"Beyond Here and Now"* feature also requires a theory of mind; one function of belief, it was pointed out, is to transmit causality to other times and places. *Hierarchical structure* first appears with composite event structure. Once there is protolanguage, in the sense in which I am using the term, there is a *lexicon* in the true sense. The significance of *temporal order* of elements in a message begins somewhere between the development of protolanguage and real syntax. *Learnability* is there from start to finish, in every stage.

Causal associations are possible from at least the earliest stages of multicellular life. A leech that moves up a heat gradient and attempts to bite when it encounters an object is responding to a causal regularity in the world. Of course, it does not know that it is responding to causal regularities; that would require a theory of mind. But the causal associations themselves are very early. The naming that this capability enables is quite within the capability of parrots, for example. Thus, in Figure 2, we can say that causal association is pre-mammalian.

At what point are animals aware of different types of the same token? At what point do they behave as if their knowledge is encoded in a way that involves variables that can have multiple instantiations? That is, at what point are they first-order? My purely speculative guess would be that it happens early in mammalian evolution. Reptiles and birds have an automaton-like quality associated with propositional representations, but most mammals that I am at all familiar with, across a wide range of genera, exhibit a flexibility of behavior that would require different responses to different tokens of the same type. Jackendoff (1999) points out that in the ape language-training experiments, the animals are able to distinguish between "symbols for individuals (proper names) and symbols for categories (common nouns)" (p. 273), an ability that would seem to require something like variable binding.

One reason to be excited about the discovery of the mirror neuron system (Rizzolatti and Arbib, 1998) is that it is evidence of an internal representation "language" that abstracts away from a concept's role in perception or action, and thus is possibly an early solid indication of "first-order" features in the evolution of the brain.

Gestural communication, composite event structure, and a theory of mind probably appear somewhere between the separation of great apes and monkeys, and the first hominids, between 15 and 5 million years ago. Arbib discusses the recognition and repetition of composite events. There are numerous studies of the gestural communication that the great apes can perform. The evolution of the theory of mind is very controversial (e.g., Heyes, 1998), but it has certainly been argued that chimpanzees have some form of a theory of mind.

These three features can thus probably be assigned to the pre-hominid era. My, again purely speculative, guess would be that vocal communication (beyond alarm cries) emerged with Homo erectus, and I would furthermore guess that they were capable of protolanguage – that is, stringing together a few words or signals to convey novel though not very complex messages. It's impossible to say how large their lexicon would have been, although it might be possible to estimate on the basis of their life style. Finally, fully modern language probably emerged simultaneously with Homo sapiens sapiens, and is what gave us a competitive advantage over our hominid cousins. We were able to construct more complex messages and therefore were able to carry out more complex joint action. As Dunbar (1996) has argued, fully modern language would have allowed us to maintain much larger social groups, a distinct evolutionary advantage.

A word on the adaptiveness of language: I have heard people debate whether language for hunting or language for social networks came first, and provided the impetus for language evolution. (We can think of these positions as the Mars and Venus theories.) This is a granularity mismatch. Language capabilities evolved over hundreds or thousands of generations, whereas hunting and currying social networks are daily activities. It thus seems highly implausible that there was a time when some form of language precursor was used for hunting but not for social networking, or vice versa. The obvious truth is that language is for establishing mutual belief, enabling joint action, and that would be a distinct advantage for both hunting and for building social networks.

4.2 A Holophrastic Stage?

Wray (1998) proposes a picture of one stage of the evolution of language that is somewhat at odds with the position I espouse in Section 3, and it therefore merits examination here. She argues that there was a holophrastic stage in the evolution of language. First there were utterances – call them protowords – that denoted situations but were not broken down into words as we know them. These protowords became more and more complex as the lexicon expanded, and they described more and more complex situations. This is the holophrastic stage. Then these protowords were analyzed into parts, which became the constituents of phrases. One of her examples is this: Suppose by chance "mebita" is the protoword for "give her the food", and "kameti" is the protoword for "give her the stone". The occurrence of "me" in both is noted and is then taken to represent a singular female recipient.

Jackendoff (1999) points out that one of the important advances leading to language was the analysis of words into individual syllables and then into individual phonemes, providing an inventory out of which new words can be constructed. It is very likely that this happened by some sort of holophrastic process. We first have unanalyzed utterances "pig" and "pit" and we then analyze them into the sounds of p, i, g, and t, and realize that further words can be built out of these elements. This process, however, is much more plausible as an account of the evolution of phonology than it is of the evolution of syntax. The phonological system is much simpler, having many fewer elements, and phonemes have no semantics to overconstrain decompositions, as words do.

Wray says, "There is a world of difference between picking out the odd word, and forcing an entire inventory of arbitrary phonetic sequences representing utterances through a complete and successful analysis." (p. 57) Indeed there is a world of difference. The latter problem is massively overconstrained, and a solution is probably mathematically impossible. This is true even if the requirement of "complete and successful" is relaxed somewhat, as Wray goes on to do. The only way I could imagine such a development would be if the individuals were generating the protowords according to some implicit morphology, and the analysis was in fact a discovery of this morphology. If children go through such a process, this is the reason it is possible. They are discovering the syntax of adult language.

A holophrastic stage has sometimes been hypothesized in child language. Children go through a one-word stage followed by a two-word stage. The holophrastic stage would be between these two. The evidence is from "words" like "allgone", "whazzat", and "gimme". An alternative explanation for these holophrases is that the child has failed to segment the string, due to insufficient segmentation ability, insufficient vocabulary, insufficient contrastive data, and so on. For a holophrastic stage to exist, we would have to show that such holophrases don't occur in the one-word stage, and I know of no evidence in support of this.

In any case, children have models whose language is substantially in advance of their own. That was never the case in language evolution. Holophrasis in child language is a misanalysis. There was nothing for holophrasis in language evolution to be a misanalysis of.

A possible interpretation of Wray's position is that originally, in evolution and in development, protowords only describe situations. Thus, a baby's "milk" might always describe the situation "I want milk." At a later stage, situations are analyzed into objects and the actions performed on them; language is analyzed into its referential and predicational functions; the lexicon is analyzed into nouns and verbs. This then makes possible the two-word stage. I take Arbib (Chapter 1, this volume) to be arguing for something like this position. I do not find this implausible, although the evidence for it is unclear. The (controversial) predominance of nouns labelling objects in children's one-word stage would seem a counterindication, but perhaps those nouns originally denote situations for the child. But I read Wray as saying there is a further analysis of protowords describing situations into their protoword parts describing objects and actions, and this seems to me quite implausible for the reasons stated.

I believe the coherence structure of discourse (e.g., Hobbs, 1985) provides a more compelling account of the evolution of the sentence. Discourse and interaction precede language. Exchanges and other reciprocal behavior can be viewed as a kind of protodiscourse. Events in the world and in discourse cohere because they stand in coherence relations with each other. Among the relations are causality:

"Smoke. Fire."

similarity:

I signal that I go around to the right. I signal that you go around to the left.

ground-figure:

"Bushes. Tiger."

occasion, or the next step in the process:

You hand me grain. I grind it.

"Approach antelope. Throw spear."

"Scalpel. Sponge."[8]

and the predicate-argument or argument-predicate relation:

"Sock. On."

"Antelope. Kill."

I point to myself. I point to the right.

While the evidence for a holophrastic stage in children's language development is scant, there *is* a stage that does often precede the two-word stage. Scollon, (1979) and others have noted the existence of what have been called "vertical constructions". Children convey a two-concept message by successive one-word utterances, each with sentence intonation, and often with some time and some interaction between them. Hoff (2001, p. 210) quotes a child near the end of the one-word stage saying, "Ow. Eye." Scollon reports a similar sequence: "Car. Go." In both of these examples, the adjacency conveys a predicate-argument relation.

It seems much more likely to me that the road to syntax was via coherence relations between successive one-word utterances, as described in Section 3, rather than via holophrasis. The holophrastic account requires some mysterious implicit derivational morphology that just happened to underlie lexical items. The coherence account requires no new mechanisms. It is just a matter of adding constraints on the interpretation of temporal order as indicating predicate-argument relations. Construction is more plausible than deconstruction.

I think Wray exaggerates the importance of grammar in communication. She says, "Successful linguistic *comprehension* requires grammar, even if the *production* were to be grammarless. A language that lacks sufficient lexical items and grammatical relations can only hint at explicit meaning, once more than one word at a time is involved. " (pp. 48-49) The problem with this statement is that discourse today has no strict syntax of the sort that a

---

[8] To pick a modern example.

sentence has, and we do just fine in comprehending it. In a sense, discourse is still in the protolanguage stage. The adjacency of segments in discourse tells hearers to figure out a relation between the segments, and normally hearers do, using what they know of context.

Context has always been central in communication. The earliest utterances were one more bit of information added to the mass of information available in the environment. In the earliest discourse, understanding the relation between utterances was part of arriving at a coherent picture of the environment. The power of syntax in modern language, as Wray points out, is to constrain interpretations and thereby lessen the burden placed on context for interpretation and to enable the construction of more complex messages, culminating in communicative artifacts cut free from physical copresence and conveying very complex messages indeed, such as this book. But there was never a point at which situations involving more than one communicative act would have been uninterpretable.

Bickerton (2003) gives further persuasive arguments against a holiphrastic stage in language evolution.

A succinct though perhaps crude formulation of my position is that it is more plausible that "Lion attack." derived from "Lion. Attack." than from "Lionattack."

4.3 Language or Language Readiness?

Arbib (Chapter 1, this volume) expresses his belief that the first physically modern Homo sapiens sapiens did not have language, only language readiness. This is a not uncommon opinion. In most such accounts, language is a cultural development that happened with the appearance of preserved symbolic artifacts, and the date one most often hears is around thirty-five to seventy thousand years ago. In one possible version of this account, anatomically modern humans of 150,000 years ago were language ready, but they did not yet have language. Language was a cultural achievement over the next 100,000 years, that somehow coincided with the species' spread over the globe.

Davidson (2003) presents a careful and sophisticated version of this argument. He argues, or at least suggests, that symbols are necessary before syntax can evolve, that surviving symbolic artifacts are the best evidence of a capacity for symbolism, and that there is no good evidence for symbolic artifacts or other symbolic behavior before 70,000 years ago in Africa and 40,000 years ago elsewhere, nor for symbolic behavior in any species other than Homo sapiens sapiens. (For example, he debunks reports of burials among Neanderthals.)

Although Davidson is careful about drawing it, the implication is if Homo sapiens sapiens evolved around 200,000 years ago and did not engage in symbolic behavior until 70,000 years ago, and if language is subsequent to that, then language must be a cultural rather than a biological development. (However, Davidson also casts doubt on the assignment of fossils to species and on the idea that we can tell very much about cognition from fossils.)

One problem with such arguments is that they are one bit of graffiti away from refutation. The discovery of one symbolic artifact could push our estimates of the origin of symbolic behavior substantially closer to the appearance of Homo sapiens sapiens, or before. Barber and Peters (1992) gave 40,000 to 35,000 years ago as the date at which humans had to have had syntax, on the basis of symbolic artifacts found up to that point. Davidson pushes that back to 70,000 years ago because of ochre found recently at a South African site and presumed to be used for bodily decoration. There have been a spate of recent discoveries of possible artifacts with possible symbolic significance. Two ochre plaques engraved with a criss-cross pattern, with no apparent nonsymbolic utility, dated to 75,000 years ago, was found at Blombos Cave in South Africa (Henshilwood et al., 2002). Pierced shells claimed to have been used as beads and dated to 75,000 years ago were found at the same site (Henshilwood et al., 2004). Rocks stained with red ochre and believed to be used in burial practices were found in Qafzeh Cave in Israel (Hovers et al., 2003); they were dated to 100,000 years ago. In northern Spain a single finely crafted pink stone axe was found in association with the fossilized bones of 27 Homo heidelbergensis individuals and is claimed as evidence for funeral rites; this site dates to 350,000 years ago (Carbonell et al., 2003). A 400,000-year-old stone object which is claimed to have been sculpted into a crude human figurine was found in 1999 near the town of Tan-Tan in Morocco (Bednarik, 2003). All of these finds are controversial, and the older the objects are purported to be, the more controversial they are. Nevertheless, they illustrate the perils of drawing conclusions about language evolution from the surviving symbolic artifacts that we so far have found.

The reason for attempting to draw conclusions about language from symbolic artifacts is that that (along with skull size and shape) is the only archaeological evidence that is remotely relevant. However, I believe it is only remotely relevant. Homo sapiens sapiens could have had language for a long time before producing symbolic artifacts. After all, children have language for a long long time before they are able to produce objects capable of lasting for tens of thousands of years. We know well that seemingly simple achievements are hard won. Corresponding to Arbib's concept of *language* readiness, we may hypothesize something called *culture* readiness. Symbolic material culture with some permanence may not have happened until 75,000 years ago, but from the beginning of our species we had *culture* readiness. The most reasonable position is that language is not identical to symbolic culture. Rather it is a component of culture readiness. As Bickerton (2003) puts it, "syntacticized language enables but it does not compel." (p. 92)

One reservation should be stated here. It is possible that non-African humans today are not descendents of the Homo sapiens sapiens who occupied the Middle East 100,000 to 90,000 years ago. It is possible rather that some subsequent stress, such as glaciation or a massive volcanic eruption, created a demographic bottleneck that would enable further biological evolution, yielding an anotomically similar Homo sapiens sapiens, who however now had fully modern cognitive capacities, and that today's human population is all descended from that group. In that case, we would have to move the date for fully modern language forward, but the basic features of modern language would still be a biological rather than a cultural achievement.

I think the strongest argument for the position that fully modern language, rather than mere language readiness, was already in the possession of the earliest Homo sapiens sapiens comes from language universals. In some scholarly communities it is fashionable to emphasize how few language universals there are; Tomasello (2003), for example, begins his argument for the cultural evolution of language by emphasizing the diversity of languages and minimizing their common core. In other communities the opposite is the case; followers of Chomsky (e.g., 1975, 1981), for example, take it as one of the principal tasks of linguistics to elucidate Universal Grammar, that biologically-based linguistic capability all modern humans have, including some very specific principles and constraints. Regardless of these differing perspectives, it is undeniable that the following features of language, among others, are universal:

- All languages encode predicate-argument relations and assertion-modification distinctions by means of word order and/or particles/inflection.
- All languages have verbs, nouns, and other words.
- All languages can convey multiple propositions in single clauses, some referential and some assertional.
- All languages have relative clauses (or other subordinate constructions that can function as relative clauses).
- Many words have associated, grammatically realized nuances of meaning, like tense, aspect, number, and gender, and in every language verbs are the most highly developed in this regard, followed by nouns, followed by the other words.
- All languages have anaphoric expressions.

These universal features of language may seem inevitable to us, but we know from formal language theory and logic that information can be conveyed in a very wide variety of ways. After the African/non-African split 100,000 to 90,000 years ago, uniform diffusion of features of language would have been impossible. It is unlikely that distant groups not in contact would have evolved language in precisely the same way. That means that the language universals were almost surely characteristic of the language of early Homo sapiens sapiens, before the African/non-African split.

It may seem as if there are wildly different ways of realizing, for example, relative clauses. But from Comrie (1981) we can see that there are basically two types of relative clause – those that are adjacent to their heads and those that replace their heads (the internal-head type). The approach of Section 3.4 handles both with minor modifications of axioms using the same predicate *Syn*; at a deep level both types pose the problem of indicating what the head is and what role it plays in the relative clause, and the solutions rely on the same underlying machinery. In any case, there is no geographical coherence to the distribution of these two types that one would expect if relative clauses were a cultural development.

There are some features of language that may indeed be a cultural development. These are features that, though widespread, are not universal, and tend to exhibit aereal patterns. For example, I would be prepared to believe that such phenomena as gender, shape classifiers, and definiteness developed subsequently to the basic features of language, although I know of no evidence either way on this issue.

There are also areas of language that are quite clearly relatively recent cultural inventions. These include the grammar of numbers, of clock and calendar terms, and of personal names, and the language of mathematics. These tend to have a very different character than we see in the older parts of language; they tend to be of a simpler, more regular structure.

If language were more recent than the African/non-African split, we would expect to see a great many features that only African languages have and a great many features that only non-African languages have. If, for example, only African languages had relative clauses, or if all African languages were VSO while all non-African languages were SVO, then we could argue that they must have evolved separately, and more recently than 90,000 years ago. But in fact nothing of the sort is the case. There are very few phenomena that occur only in African languages, and they are not widespread even in Africa, and are rather peripheral features of language; among these very few features are clicks in the phonology and logophoric pronouns, i.e., special forms of pronouns in complements to cognitive verbs that refer to the cognizer. There are also very few features that occur only in non-African languages. Object-initial word order is one of them. These features are also not very widespread.[9]

Finally, if language were a cultural achievement within the last 50,000 years, rather than a biological achievement, we would expect to see significant developments in language in the era that we have more immediate access to, the last five or ten thousand years. For example, it might be that languages were becoming more efficient, more learnable, or more expressive in historical times. As a native English speaker, I might cite a trend from inflection and case markings to encode predicate-argument relations to word order for the same purpose. But in fact linguists detect no such trend. Moreover, we would expect to observe some unevenness in how advanced the various languages of the world are, as is the case with technology. Within the last century there have been numerous discoveries of relatively isolated groups with a more primitive material culture than ours. There have been no discoveries of isolated groups with a more primitive language. All of these arguments should lead us to reject the possibility that the basic features of language are a cultural acquisition subsequent to the appearance and dispersion of Homo sapiens sapiens. On the contrary, fully modern language has very likely been, more than anything else, what made us human right from the beginning of the history of our species.

Acknowledgments: This chapter is an expansion of a talk I gave at the Meeting of the Language Origins Society in Berkeley, California, in July 1994. The original key ideas arose out of discussions I had with Jon Oberlander, Mark Johnson, Megumi Kameyama, and Ivan Sag. I have profited more recently from discussions with Lokendra Shastri, Chris Culy, Cynthia Hagstrom, and Srini Narayanan, and with Michael Arbib, Dani Byrd, Andrew Gordon, and the other members of Michael Arbib's language evolution study group. Michael Arbib's comments on the original draft of this chapter have been especially valuable in strengthening its arguments. None of these people would necessarily agree with anything I have said.

References

Akmajian, Adrian, and Chisato Kitagawa, 1974. "Pronominalization, Relativization, and Thematization: Interrelated Systems of Coreference in Japanese and English", Indiana University Linguistics Club.

---

[9] I have profited from discussions with Chris Culy on the material in this paragraph.

Barber, E. J. W., and A. M. W. Peters, 1992. "Ontogeny and Phylogeny: What Child Language and Archaeology Have to Say to Each Other". In J. A. Hawkins and M. Gell-Mann (Eds.), *The Evolution of Human Languages*, Addison-Wesley Publishing Company, Reading, Massachusetts, pp. 305-352.

Bednarik, Robert G., 2003. "A Figurine from the African Acheulian", *Current Anthropology*, Vol. 44, No. 3, pp. 405-412.

Bickerton, Derek, 1990. *Language and Species*, University of Chicago Press, Chicago.

Bickerton, Derek, 2003. "Symbol and Structure: A Comprehensive Framework for Language Evolution", in M. H. Christiansen and S. Kirby (Eds.), *Language Evolution*, Oxford University Press, Oxford, United Kingdom, pp. 77-93.

Carbonell, Eudald, Marina Mosquera, Andreu Ollé, Xosé Pedro Rodriguez, Robert Sala, Josep Maria Vergès, Juan Luis Arsuaga, and José María Bermúdez de Castro, 2003. "Les premier comportements funéraires auraient-ils pris place à Atapuerca, il y a 350 000 ans?" *L'Anthropologie*, Vol. 107, pp. 1-14.

Chomsky, Noam, 1975. *Reflections on Language*, Pantheon Books, New York.

Chomsky, Noam, 1981. *Lectures on Government and Binding*, Foris, Dordrecht, Netherlands.

Chomsky, Noam, 1995. *The Minimalist Program*, MIT Press, Cambridge, Massachusetts.

Clark, Herbert, 1975. "Bridging". In R. Schank and B. Nash-Webber (Eds.), *Theoretical Issues in Natural Language Processing*, pp. 169-174. Cambridge, Massachusetts.

Comrie, Bernard, 1981. *Language Universals and Linguistic Typology*, University of Chicago Press, Chicago.

Davidson, Iain, 2003. "The Archaeological Evidence for Language Origins: States of Art", in M. H. Christiansen and S. Kirby (Eds.), *Language Evolution*, Oxford University Press, Oxford, United Kingdom, pp. 140-157.

Dunbar, Robin, 1996. *Grooming, Gossip and the Evolution of Language*. Faber and Faber, London.

Engel, Andreas K., and Wolf Singer, 2001. "Temporal Binding and the Neural Correlates of Sensory Awareness", *Trends in Cognitive Science*, Vol. 5, pp. 16-25.

Fell, Jürgen, Peter Klaver, Klaus Lehnertz, Thomas Grunwald, Carlo Schaller, Christian E. Elger, and Guillén Fernandez, 2001. "Human Memory Formation is Accompanied by Rhinal-Hippocampal Coupling and Decoupling", *Nature Neuroscience*, Vol. 4, pp. 1259-1264.

Fikes, Richard, and Nils J. Nilsson, 1971. "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving", *Artificial Intelligence*, Vol. 2, pp. 189-208.

Grice, Paul, 1948. "Meaning", in *Studies in the Way of Words*, Harvard University Press, Cambridge, Massachusetts, 1989.

Henshilwood, Christopher, Francesco d'Errico, Marian Vanhaeren, Karen van Niekirk, and Zenobia Jacobs, 2004. "Middle Stone Age Shell Beads from South Africa", *Science*, Vol. 304, p. 404.

Henshilwood, Christopher, Francesco d'Errico, Royden Yates, Zenobia Jacobs, Chantal Tribolo, Geoff A. T. Duller, Norbert Mercier, Judith C. Sealy, Helene Valladas, Ian Watts, and Ann G. Wintle, 2002. "Emergence of Modern Human Behavior: Middle Stone Age Engravings from South Africa", *Science*, Vol. 295, pp. 1278-1280.

Heyes, Cecilia M., 1998. "Theory of Mind in Nonhuman Primates", *Behavioral and Brain Sciences*, Vol. 21, pp. 101-148.

Hobbs, Jerry R., 1985a. "Ontological Promiscuity", *Proceedings*, 25th Annual Meeting of the Association for Computational Linguistics, Chicago, Illinois, July 1985, pp. 61-69.

Hobbs, Jerry R., 1985b. "On the Coherence and Structure of Discourse", Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Hobbs, Jerry R. 1998. "The Syntax of English in an Abductive Framework", Available at http://www.isi.edu/hobbs/discourse-inference/chapter4.pdf

Hobbs, Jerry R., 2001. "Syntax and Metonymy", in P. Bouillon and F. Busa (Eds.), *The Language of Word Meaning*, Cambridge University Press, Cambridge, United Kingdom, pp. 290-311.

Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. "Interpretation as Abduction", *Artificial Intelligence*, Vol. 63, Nos. 1-2, pp. 69-142.

Hoff, Erika, 2001. *Language Development*, Wadsworth, Belmont, California.

Hovers, Erella, Shimon Ilani, Ofer Bar-Yosef, and Bernard Vandermeersch, 2003. "An Early Case of Color Symbolism: Ochre Use by Modern Humans in Qafzeh Cave", *Current Anthropology*, Vol. 44, No. 4, p. 491-522.

Jackendoff, Ray, 1999. "Possible Stages in the Evolution of the Language Capacity", *Trends in Cognitive Sciences*, Vol. 3, No. 7, pp. 272-279.

Kameyama, Megumi, 1994. "The Syntax and Semantics of the Japanese Language Engine", in R. Mazuka and N. Nagai, eds. *Japanese Syntactic Processing*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Klein, Wolfgang, and Clive Perdue, 1997. "The Basic Variety, or Couldn't Language Be Much Simpler?", *Second Language Research*, Vol. 13, pp. 301-347.

Pierce, Charles S., 1955 [1903]. "Abduction and Induction", in J. Buchler (Ed.), *Philosophical Writings of Pierce*, Dover Publications, New York, pp. 150-156.

Pollard, Carl, and Ivan A. Sag, 1994. *Head-Driven Phrase Structure Grammar*, University of Chicago Press, Chicago, and CSLI Publications, Stanford, California.

Premack, David, and Guy Woodruff, 1978. "Does the Chimpanzee Have a Theory of Mind?" *Behavioral and Brain Sciences*, Vol. 1, No. 4, pp. 515-526.

Rizzolati, Giacomo, and Michael A. Arbib, 1998. "Language Within Our Grasp", *Trends in Neurosciences*, Vol. 21, No. 5, pp. 188-194.

Scollon, Ronald, 1979. "A Real Early Stage: An Unzippered Condensation of a Dissertation on Child Language", in Elinor Ochs and Bambi B. Schiefielin (Eds.), *Developmental Pragmatics*, Academic Press, New York, pp. 215-227.

Shastri, Lokendra, 1999. "Advances in SHRUTI – A Neurally Motivated Model of Relational Knowledge Representation and Rapid Inference Using Temporal Synchrony", *Applied Intelligence*, Vol. 11, pp. 79-108.

Shastri, Lokendra, 2001. "Biological Grounding of Recruitment Learning and Vicinal Algorithms in Long-Term Potentiation", in J. Austin, S. Wermter, and D. Wilshaw (Eds.), *Emergent Neural Computational Architectures Based on Neuroscience*, Springer-Verlag, Berlin.

Shastri, Lokendra, and Venkat Ajjanagadde, 1993. "From Simple Associations to Systematic Reasoning: A Connectionist Representation of Rules, Variables and Dynamic Bindings Using Temporal Synchrony", *Behavioral and Brain Sciences*, Vol. 16, pp. 417-494.

Shastri, Lokendra, and Carter Wendelken, 2003. "Learning Structured Representations", *Neurocomputing*, Vol. 52-54, pp. 363-370.

Wason, P. C., and Philip Johnson-Laird, 1972. *Psychology of Reasoning: Structure and Content*, Harvard University Press, Cambridge, MA.

Wendelken, Carter, and Lokendra Shastri, 2003. "Acquisition of Concepts and Causal Rules in SHRUTI", *Proceedings*, Twenty-Fifth Annual Conference of the Cognitive Science Society.

Wray, Alison, 1998. “Protolanguage as a Holistic System for Social Interaction”, *Language and Communication*, Vol. 18, pp. 47-67.